Retroflex-Triggered “Sound Swallowing” in Beijing Mandarin: Acoustic Properties and a Preliminary Discussion on its Phonologization

Yuying Zhu 14452685
BA Thesis Linguistics (extended for the Honours programme)
Supervised by Dr. A.T. Benders
University of Amsterdam
23 June 2025

Abstract

"Sound Swallowing (*Tūnyīn*, *吞音*, hereafter *Beijing Swallowing*)" is an understudied phenomenon in Beijing Mandarin. Of all processes characterized as "Beijing Swallowing", "swallowing" elicited by retroflex obstruents was most frequently and consistently reported in prior studies (e.g., Han, 2024; Chirkova & Chen, 2012; Zhang, 2005). That is, the syllable rime preceding the retroflex takes on a final [ɻ], and the syllable containing the swallowing-triggering retroflex gets fully dropped, for instance, pu^{51} ʈ͡ʂə^{55} tɑʊ^{51} → puɻ^{51} tɑʊ^{51} (*不知道*, 'do not know'). However, this claim is supported by limited acoustic evidence.

The current study aims to investigate the acoustic features of "Beijing Swallowing" triggered by retroflex obstruents. 15 native Beijing speakers participated in a trisyllabic sequence production task, producing a given list of stimulus sequences under different conditions ("unswallowed", "swallowed" and "rhotacized"). The results suggest that retroflex-triggered "swallowing" involves merging the syllable that contains the retroflex and its preceding syllable into one and reducing their overall duration, while lowering the average F3 value of the rime preceding the retroflex. Furthermore, results suggest that retroflex "swallowing" and [ɻ] suffixation should be treated as two distinct processes on the phonetic level.

A sentence production task was also conducted to examine the relationship between speech rate ("normal", "slow" and "very slow") and the presence of retroflex "swallowing", in order to preliminarily discuss the potential phonologization of this process. The results indicate that, although the frequency of "swallowing" is still sensitive to speech rate, acoustic features of retroflex "swallowing" can be observed in slow, and even occasionally in very slow speech. This suggests that retroflex-induced "swallowing" might be undergoing phonologization, challenging the previous claim by Han (2024) that the process is fully phonetic.

Acknowledgements

First, I would like to express my sincere gratitude to my supervisor, Dr. A.T. Benders, for her invaluable feedback, guidance, and support throughout the whole thesis journey. I would also like to thank technician Dirk J. Vet at the UvA Speech Lab for his assistance with recording equipment and *Praat* scripting. Special thanks to everyone who helped with recruitment and all my participants for their enthusiasm, this study wouldn't have been possible without your aid. Thanks to everyone who generously shared and discussed their code on *StackOverflow*, *CSDN*, *Zhihu*, *YouTube*, and many other sources I cannot name one by one, from whom I learned greatly. Thanks to my friends Zehua and Chunying for their companion and mental support. Lastly, I would like to thank my parents for their unyielding support in every possible way. My gratitude to them is beyond words.

Table of Contents

1. Theoretical Background *6*

1.1. Beijing Mandarin **6**

1.2. "Sound Swallowing" in Beijing Mandarin **6**

1.3. "Sound Swallowing" in Tianjin Mandarin **9**

1.4. Limitations of previous studies **10**

2. The Current Study *12*

2.1. Aims **13**

2.2. Why retroflexes? **13**

2.3. How to determine whether something is phonetic or phonological? **14**

2.4. Research questions and approaches **15**

3. Methods *15*

3.1. Participants **15**

3.2. Stimuli **17**

3.2.1. Part 1: Trisyllabic production task 17

3.2.2. Part 2: Sentence production task 18

3.3. Experimental design **18**

3.3.1. Part 1: Trisyllabic production task 18

3.3.2. Part 2: Sentence production task 20

3.4. Procedure **21**

3.5. Data processing & measurements **22**

3.5.1. Pre-processing 22

3.5.2. Data processing & measurements 22

3.5.2.1. Trisyllabic production task 22

3.5.2.2. Sentence production task 24

4. Results *24*

4.1. Part1: Trisyllabic production task **25**

4.1.1. Number of syllable(s) in non-final part of recorded utterances 25

4.1.2. Durations of recorded utterances' non-final part 26

4.1.3. F3 values 28

4.2. Part 2: Sentence production task **38**

4.2.1. Number of syllable(s) in the non-final part of the targeted location names 38

4.2.2. Durations of the non-final part of the targeted location names 40

5. Discussion *42*

5.1. Summary of results and answering the research questions **42**

5.2. Limitations **43**

5.3. Is there actually phonologization going on? **45**

5.4. Suggestions for future study **46**

6. Conclusion *47*
References *48*
Appendix 1: Demographic profile of participants *51*
Appendix 2: Stimuli *52*
Appendix 3: Descriptive statistics on average F3 values in initial syllable's rime *54*
Appendix 4: Information brochure and consent form (with English translation) *56*

1. Theoretical Background

1.1. Beijing Mandarin

Beijing Mandarin (*Běijīnghuà*, *北京话*) is a northern dialect of Mandarin Chinese, spoken in urban Beijing by its residents (Chirkova & Chen, 2012). Beijing Mandarin holds a unique place among all contemporary Mandarin variants, as it serves as the phonological base of modern Standard Mandarin (*Pǔtōnghuà, 普通话*). The latter is the official language of Mainland China and a lingua franca used by Chinese-speaking communities worldwide (Chirkova & Chen, 2012; Handel, 2017; Li, 2006).

Despite its prominent status, documentation and academic studies on Beijing Mandarin are limited, especially in its phonetics. This could be explained by several factors identified in previous studies: First, Beijing Mandarin is often overlooked as a distinct subject of study from Standard Mandarin, as the two are conventionally considered as phonologically comparable. However, previous studies have found Beijing Mandarin to differ considerably in all linguistic sub-systems from Standard Mandarin, even in its phonological organization (Astraxan et al., 1985; Zhū, 1987; Lín et al., 1987; as cited in Chirkova & Chen, 2012). As a result, while Standard Mandarin is the most researched Mandarin variant, its studies are not fully generalizable to Beijing Mandarin. Second, according to Chirkova & Chen (2012), Beijing Mandarin's underrepresentation is also related to its lower prestige, associated with the lower education and socio-economic status of its speakers. Moreover, as stated in Ingebretson (2025), the language policy in Mainland China promoting Standard Mandarin caused a loss of other Chinese variants, which might also have contributed to the neglect of Beijing Mandarin in academic research.

1.2. "Sound Swallowing" in Beijing Mandarin

Besides its rich colloquial lexicon (e.g., Zhang, 2014) and extensive use of "er-hua" rimes (i.e., syllable-final rhoticity, e.g., Lee, 2005; Xing, 2021), another feature distinguishing Beijing Mandarin from Standard Mandarin is a phenomenon known as "Sound Swallowing (*Tūnyīn*, *吞音*, hereafter *Beijing Swallowing*, or simply, *swallowing*. See (1)-(3) for examples)".

There are only very limited previous studies on “Beijing Swallowing”. Consequently, the linguistic nature of this phenomenon remains largely unclear. Several studies discussed the relationship between “Beijing Swallowing” and speech prosody. Specifically, trisyllabic Mandarin sequences often exhibit a “medium–weak–strong” prosodic structure, and the middle syllable is most sensitive to “swallowing” (Chao, 2005; Yan & Lin, 1988; Wang & Wang, 1993; as cited in Han, 2024). “Beijing Swallowing” was also briefly mentioned in other studies that focused on the sociolinguistic aspect of Beijing Mandarin. For instance, Zhang (2005, 2021) described “swallowing” triggered by the three Mandarin retroflex obstruents (/t͡ʂ/, /t͡ʂʰ/ and /ʂ/) as a process of lenition, in which the retroflex onsets in weak syllables merge into [ɻ], resulting in the rhotacization of the rime of the syllable preceding the retroflex. This pattern was also reported by other studies, as shown in examples (1)-(3).

(1) Swallowing triggered by yoiceless retroflex fricative /ʂ/ (Chirkova & Chen, 2012):

tjɛn^{51} ʂə̯^{51} tʰaɪ^{35} → tiɛ̃ɻ^{51}tʰai^{35} (*电视台*, ‘TV station’)

(2) Swallowing triggered by aspirated retroflex affricate /t͡ʂʰ/ (Zhang, 2005):

pʰai t͡ʂʰu suo → pʰaiɻ suo (*派出所*, ‘police station’)

(3) Swallowing triggered by unaspirated retroflex affricate /t͡ʂ/ (Han, 2024):

pu^{51} t͡ʂə̯^{55} tɑʊ^{51} → puɻ^{51} tɑʊ^{51} (*不知道*, ‘do not know’)

For reference, the superscripted numbers in examples (1)-(3) represent the Mandarin lexical tones in Chao tone numerals (Chao, 1930, as cited in Bao, 1990), describing the pitch countours with a five-level frequency scale (1 = lowest, 5 = highest). This transcription method will be used throughout this paper if not otherwise specified. An overview of Chao tone numerals is provided in *Table 1*.

Table 1. The four Mandarin lexical tones, described in Chao tone numerals (Chao, 1930)

Tone description	Chao tone numbers
High-level	55
Mid-rising	35
Low-dipping	214
High-falling	51

To the author's knowledge, to date, there is only one relatively comprehensive and systematic phonetic study on "Beijing Swallowing", published in 2024 by Han. Focusing on "swallowing" of native Beijing speakers in familiar trisyllabic sequences, Han identified three major "swallowing" patterns:

A. Merger of the middle and the initial syllable, wherein the trisyllabic sequences being reduced to disyllabic, e.g.:

(3) pu^{51} $\widehat{\text{ʈʂ}}ə^{55}$ $tɑʊ^{51}$ → $puɻ^{51}$ $tɑʊ^{51}$ (*不知道*, 'do not know')

B. Partial reduction of the middle syllable, e.g.:

(4) xu^{51} kuo^{35} $sɤ^{51}$ → xu^{51} uo^{35} $sɤ^{51}$ (*护国寺*, location name)

C. Complete deletion of the middle syllable with no trace left behind, e.g.:

(5) $tuŋ^{51}$ u^{51} yan^{35} → $tuŋ^{51}$ yan^{35} (*动物园*, 'zoo')

For each major "swallowing" pattern, Han also recognized several sub-patterns. According to Han, these patterns displayed significant arbitrariness, and the "swallowing" behavior recorded exhibited notable individual-, item-, and context-specific variability. Moreover, a "swallowed" form was reported to show no semantic difference from the "unswallowed" form. Based on these observations, Han concluded that "Beijing Swallowing" is a purely phonetic phenomenon[1], primarily driven by speech rate and articulatory economy.

[1] Han's reasoning, surprising as it may seem, is translated as follows: "'Beijing Swallowing' is a purely phonetic phenomenon: Unlike neutral tone or rhotacization, which, while phonetic in nature, are also closely related to semantics and syntactic structure, 'swallowing' should be considered entirely phonetic, with little to no impact on meaning […] Precisely because of these characteristics, 'swallowing' should be considered a relatively individual phonetic phenomenon."

1.3. "Sound Swallowing" in Tianjin Mandarin

Han's study was largely inspired by a previous study on "Sound Swallowing" in Tianjin Mandarin (Wee et al., 2005), another Northern Mandarin dialect spoken in a region about 120 km from Beijing[2] (Li et al., 2017). According to Han and Wee et al., "Sound Swallowing" is a largely shared feature of Beijing and Tianjin Mandarin. Therefore, the studies on "Tianjin Swallowing" (referred to by Wee et al. as "casual speech elision") might offer some insights into "Beijing Swallowing".

Wee (2008, 2014) analyzed "Tianjin Swallowing" as a process of "elide-and-merge". Specifically, they proposed that trisyllabic Mandarin sequences contain two inter-syllabic "windows" (see *Figure 1*), and phonological material, especially consonants, can be elided at the first window (Window I) between the two non-final syllables. As "swallowing" takes place, the coda of the initial syllable and the onset of the medial syllable are dropped, and the remaining segments merge to form a single syllable, as illustrated in *Figure 2*.

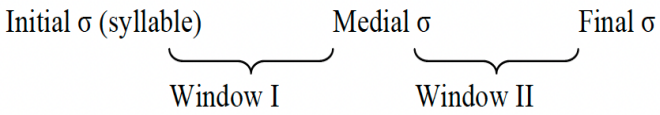

Figure 1. Wee's schematic representation of Mandarin trisyllabic strings (figure from Wee, 2014)

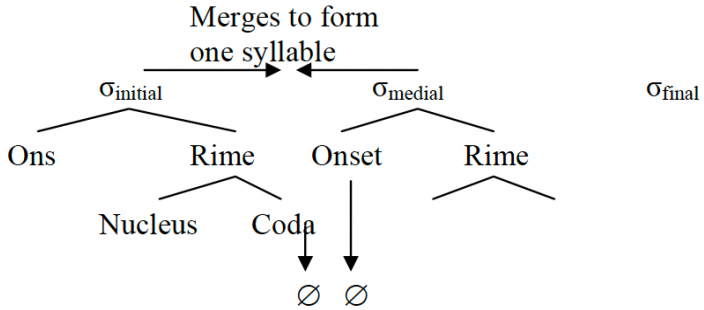

Figure 2. Wee's analysis on "Sound Swallowing" in trisyllabic Tianjin Mandarin sequences as a process of elide-and-merge (figure from Wee, 2014)

[2] Given China's vast size, i.e., 9,600,000 km^2 (https://en.wikipedia.org/wiki/Geography_of_China), the two regions are generally considered fairly near each other.

1.4. Limitations of previous studies

While offering some valuable insights into the phonetic aspect of "Beijing Swallowing", there are several limitations in the previous studies reported above:

First, Han's (2024) analysis was fully based on broad IPA transcriptions, without conducting any investigation on the acoustic level. As a result, some instances of "Beijing Swallowing" reported in the paper might be oversimplified, or even inaccurate in terms of phonetic details. For example, as shown in *Figure 3*, the pitch curve of the initial syllable in the "swallowed" trisyllabic sequence |pu^{51}.t͡ʂɚ^{55}.tɑʊ^{51}| is slightly uprising and should be described as [puɻ^{45}] or [puɻ^{55}]. However, this was inaccurately reported by Han as [puɻ^{51}] in (3), which takes the simple falling tone of the initial syllable in the "unswallowed" form. Despite potential individual variation, this might suggest that Han's report on the effect of "swallowing" on tone realization is oversimplified.

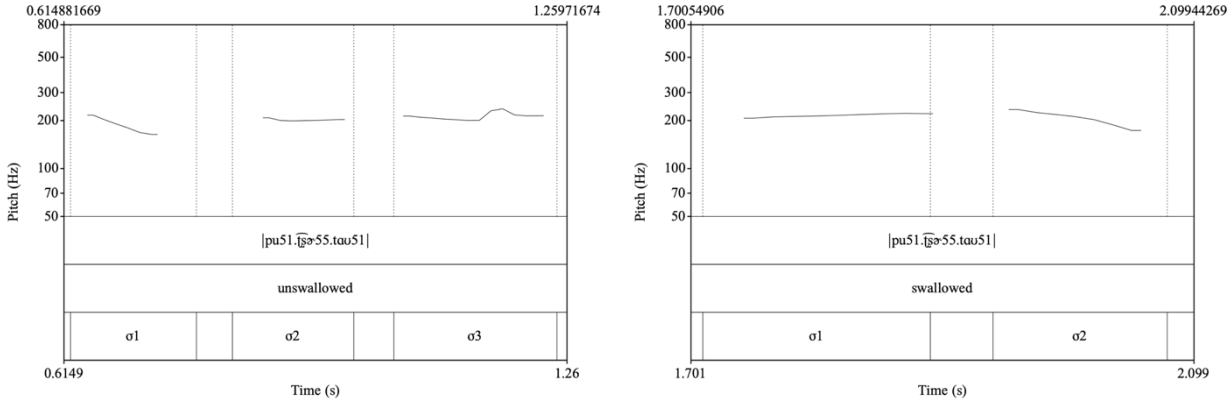

Figure 3. Pitch counters of trisyllabic sequence |pu^{51}.t͡ʂɚ^{55}.tɑʊ^{51}| from the same speaker (Participants 01 in the current study), without (left) and with (right) "swallowing"

Another example of potential inaccuracy in Han (2024) is (6), which, according to Han, is an instance of Pattern C "swallowing" (i.e., "complete deletion of the middle syllable with no trace left behind"). However, spectrographic comparison shows clear differences in the initial syllables between a "swallowed" |mu^{51}.ɕy^{55}.ti^{51}| and [mu^{51}.ti^{51}], most significantly in the trajectories of the second formant (marked with boxes in *Figure 4*). This further questions the accuracy and representativeness of Han's observations, highlighting the need to collect more reliable data for "Beijing Swallowing".

(6) mu^{51} ɕy^{55} ti^{51}→ mu^{51} ti^{51} (*木樨地*, location name)

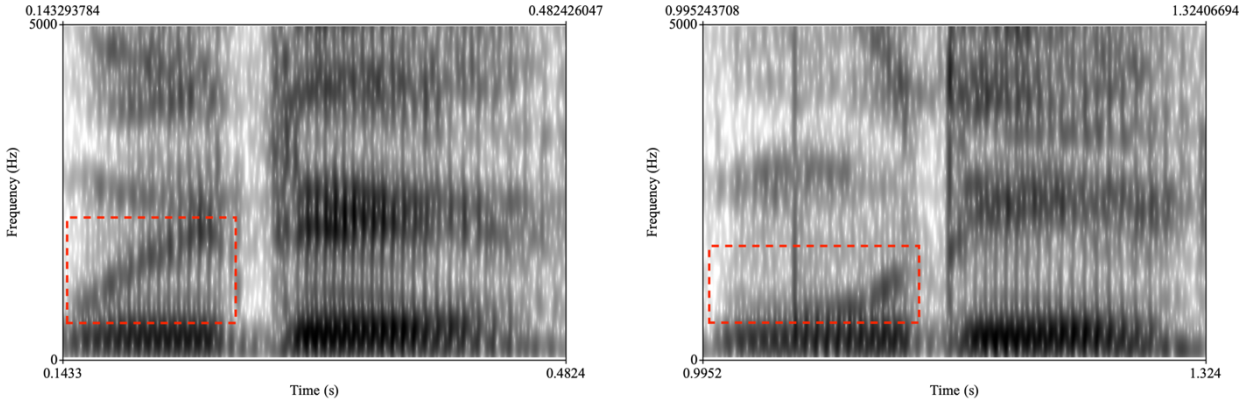

Figure 4. Spectrogram of "swallowed" trisyllabic sequence |mu^{51} ɕy^{55} ti^{51}| (left) and [mu^{51} ti^{51}] (right) from the same speaker (Participant 10 in the current study)

Second, Han's conclusion that "Beijing Swallowing" is a purely phonetic process was mostly drawn from the observation that "swallowing" had no effect on the meaning of a word for both the speaker and the listener. However, while not documented in prior research, there is one case of lexicalized "swallowing" consistently recognized by Beijing speakers. Specifically, [da˨˩^{51}.la˨˩^{35}], which is the "swallowed" form of |da^{51}.t͡ʂa^{51}.la˨˩^{35}|, is used exclusively as a location name (Liang, 2021), while the "unswallowed" form means "big fence" in the literal sense. According to some models on language change, for instance the one summarized by Hyman from various sources (2008, see example (7)), lexicalization is a subsequent process to phonologization. Although [da˨˩^{51}.la˨˩^{35}] is the only known instance of potential lexicalization of "Beijing Swallowing", it still suggests that Han's claim should be taken with caution.

(7) Hyman (2008, derived from Vennemann (1972a, b), Dressler (1976, 1985), Joseph & Janda (1988), etc.):

phonetic > phonologized > phonemicized > morphologized > lexicalized > LOSS

Furthermore, according to Han, retroflex-triggered "swallowing" resulting in the rhotacization of its preceding syllable rime was only reported to have been observed in /ʐ/ in the 1980s, (Lin, 1982; as cited in Han, 2024). An example is shown in (8). However, such "swallowing" patterns can now be found in other retroflex obstruents (see examples (1)-(3)), suggesting potential

expansion of the process in the past forty years. As this could involve the phonological level, further investigation on the phonologization of retroflex-triggered “swallowing” is required.

(8) mei^{35} zᶎən^{35} tɕʰy^{51}→ mər^{35} tɕʰy^{51} (*没人去*, ‘no one is going’)

It should also be noted that the retroflex “swallowing” patterns reported in previous studies might not be fully accurate (see examples (1)-(3)). These studies, being predominantly sociolinguistic rather than phonetic, treated “swallowed” retroflexes as akin to syllable-final rhoticity (“er-hua”) in Beijing Mandarin, transcribed with an [ɻ]. However, closer investigation of the phonetic nature of Beijing final rhoticity suggested otherwise. An acoustic study by Lee (2005) described the process as [ɚ]-suffixation. According to Lee, monophthong rhotacization results in a significant drop in the third formant (henceforth *F3*), and the rhotacized monophthongs are not retroflexed in nature. This was also confirmed by a study by Xing (2021). Xing’s study also found that vowel rhotacization in Beijing Mandarin is not realized through attaching the rhotic approximant to the vowel, but through rhoticity of the whole rime. Thus, it is suggested to test on the acoustic level, particularly with the F3 value, whether a full merger of retroflexes into [ɻ] takes place during “swallowing”.

Moreover, Han’s study landed on the conclusion that the realization of “swallowing” is highly unstable, exhibiting significant variation across speakers, contexts and items. The conditioning factors of the three major “swallowing” patterns and the many sub-patterns identified remain underspecified.

2. The Current Study

Following Han (2024) and Wee et al. (2005), the present study focuses on “Beijing Swallowing” triggered by the three Mandarin retroflex obstruents (i.e., /t͡ʂ/, /t͡ʂʰ/ and /ʂ/; /ʐ/ is not included because the realization of this sound is highly variable between speakers) in high-frequency trisyllabic sequences in native speakers of Beijing Mandarin.

2.1. Aims

The objectives of the current study are two-fold:

First, as previously noted, existing research on "Beijing Swallowing" is scarce, particularly at the acoustic level. This lack of empirical evidence significantly hinders further investigation on this phenomenon. Therefore, the primary goal of this study is to collect some reliable acoustic data on "Beijing Swallowing", providing a foundation upon which the present and future research can build. This study also aims to test the retroflex "swallowing" pattern reported in previous studies (e.g., Han, 2024; Chirkova & Chen, 2012; Zhang, 2005). That is, the syllable rime preceding the retroflex takes on a final [ɻ], while the syllable containing the swallowing-triggering retroflex gets fully dropped, for instance, pu^{51} t͡ʂɚ^{55} tɑʊ^{51} → puɻ^{51} tɑʊ^{51} (*不知道*, 'do not know').

Second, this study aims to re-examine Han's claim that "Beijing Swallowing" is a highly unstable, purely phonetic process driven by speech rate and articulatory economy, as Han's conclusions were based on limited acoustic evidence and involved notable methodological shortcomings. Namely, this study seeks to explore whether Beijing speakers' "swallowing" behavior exhibits recognizable patterns that suggest phonologization.

2.2. Why retroflexes?

Although it would be ideal to provide a comprehensive acoustic profile involving all segments that may induce "Beijing Swallowing", the current study is restricted to "swallowing" triggered by retroflex obstruents. This focus is informed by the consistent recognition of retroflex-triggered "swallowing" patterns in previous studies (Han, 2024; Zhang, 2005; Zhang, 2021; Chirkova & Chen, 2012), Han's (2024) report on the diachronic expansion of retroflex "swallowing", and the one instance of lexicalized "swallowing" in Beijing Mandarin, which also happens to involve a retroflex affricate (see *Section 1.4*). This recurrent emphasis highlights the importance of retroflex obstruents in "Beijing Swallowing" and its possible interaction with levels beyond phonetics. The author therefore considers these segments a good starting point for

investigating the complex process of “Beijing Swallowing” and its relationship with phonologization, which could inform the direction of future studies.

2.3. How to determine whether something is phonetic or phonological?

In order to re-examine the phonetic nature of “Beijing Swallowing”, some indicators for the presence or absence of phonologization, namely, the process “whereby a phonetic process becomes phonological”, are needed (Hyman, 1975, as cited in Hyman, 2008). According to Shahin (2011, summarized from various sources, e.g. Hyman, 2008), categoricality or discreteness is the only necessary indicator for phonologization. In other words, phonologization is considered to have taken place if certain phonetic forms consistently and relatively stably surface in speech. However, this is hard to incorporate into the present study due to the lack of knowledge to the acoustic characteristics associated with retroflex-triggered “swallowing”. Another indicator of phonologization proposed by Shahin is its interaction with the morphological or syntactic structures of the language, which is in line with the model summarized by Hyman (2008) in (7). Nonetheless, this is also not applicable for investigating “Beijing Swallowing”, as there is almost no report of its interactions with these levels except for the one instance of lexicalized “swallowing”.

To gather evidence for phonologization, a method based on speech rate proposed by Solé (1994) is applied in the current study. Through investigating vowel nasalization in American English and Spanish, Solé found processes with phonetic origins, like lenition or deletion, typically increase at faster speech rates. However, they can occur even in slow speech or in careful pronunciation if become phonologized. Once fully phonologized, these processes are no longer sensitive to speech rate. Another study by Solé & Ohala (2010) investigated hight-dependent vowel duration contrast in various languages. The findings suggested that the stable occurrence of a contrast across different speech rates indicate intentional control from the speaker, indicating the contrast has been phonologized. While not as necessary and conclusive as the categoricality / discreteness criteria mentioned above, the speech rate method could still serve as a useful approach for the preliminary investigation on phonologization in “Beijing Swallowing”.

2.4. Research questions and approaches

The current study consists of two main components, approaching retroflex-triggered "swallowing" from a phonetic and a phonological perspective respectively:

1) **Phonetic investigation:** This investigation aims to address the acoustic characteristics associated with retroflex-triggered "Beijing Swallowing", particularly in terms of syllable merger, duration reduction, and F3 value. The focus on F3 was informed by Lee's (2005) finding that final rhoticity in Beijing Mandarin causes F3 drop, implying potential F3 drop in retroflex "swallowing" as well. This was approached with a trisyllabic sequence production task (*trisyllabic production task* hereafter), comparing participants' production with and without "swallowing". Furthermore, the retroflex "swallowing" pattern reported in previous studies (i.e., the syllable rime preceding the retroflex takes on a final [ɪ], while the syllable containing the swallowing-triggering retroflex gets fully elided) was tested by comparing the acoustic characteristics between the "swallowed" and "rhotacized" forms, as well as comparing the effect of the vowel following the swallowing-triggering retroflex on the F3 of the initial syllable's rime in the "swallowed" forms (see *Section 3*).

2) **Phonological investigation:** This investigation aims to examine whether retroflex-triggered "Beijing Swallowing" occurs in slow speech, and how speech rate affects the frequency of "swallowing". This was approached with a sentence production task that elicits more natural-speech like production than the trisyllabic production task, in order to compare the frequency of the occurrence of acoustic characteristics associated with retroflex "swallowing" identified in the phonetic investigation across different speech rates.

3. Methods

3.1. Participants

Sixteen adult native speakers of Beijing Mandarin participated in the experiment. One participant (Participant 13, 79-year-old male) was excluded from the final analysis because he was

deemed unreliable[3]. All participants were born in urban Beijing, and were raised and educated there until at least 18 years old. They self-reported to have no physical or cognitive impairments that might affect their performance in the experimental tasks.

Recruitment was mainly conducted through *WeChat*, a Chinese instant messaging service (Tencent Holdings Limited, 2025). Recruited via a message distributed through *WeChat*, participants completed a brief questionnaire on *Qualtrics* (Qualtrics, Provo, UT) to sign up for the experiment and provide necessary information to confirm their eligibility. Participants' demographic information was collected through a questionnaire after their participation in the experiment. *Table 2* provides an overview of the participants' demographic profile. A more detailed summary is available in *Appendix 1*. This information was collected with participants' informed consent, under the approval of the *Ethics Committee of the Faculty of Humanities* at the University of Amsterdam, case number *FGW-6261*. The information brochure and consent form for the current study are provided in *Appendix 4* with English translation.

Table 2. Demographics of Participants (Participant 13 excluded)

Demographics	**Group**	**n**	**Percentage (%)**
Age (M = 28.6, max = 44, min = 18)	18 to 30 (M = 21.9, max = 25, min = 18)	11	73%
	30 or above (M = 42.0, max = 44, min = 37)	4	27%
Gender	Male	8	53%
	Female	7	47%
Family background	Two Beijing-born parents	10	67%
	One Beijing-born parent	3	20%
	No Beijing-born parent	2	13%
Education	College or above	15	100%
Knowledge in other languages aside from	One other language	12	80%
	More than one other language	3	20%

[3]The decision of exclusion was made carefully based on the following reasons: 1. The participant only provided his personal information verbally prior to the experiment, declaring that he had no physical or cognitive impairments that might affect his performance in the tasks. However, he informed the experimenter that he was wearing hearing aids halfway through the recording. 2. There was a question in the questionnaire that the participant filled out with the experimenter, asking whether he had any knowledge to languages other than Mandarin. The participant claimed to have no knowledge to any second language when filling out the questionnaire. However, in a later conversation, he disclosed that he speaks both English and Russian. 3. Although being asked to avoid verbal conversation with the experimenter during the recording unless necessary, the participant made a lot of verbal comments, making data processing particularly challenging. 4. The participant showed limited understanding to the experimental conditions used in the tasks. Given these, the participant was considered unreliable and excluded from the analysis.

Standard Mandarin and Beijing Mandarin			
Knowledge in other Mandarin variants aside from Standard Mandarin and Beijing Mandarin	None	12	80%
	One other variant	3	20%

3.2. Stimuli

3.2.1. Part 1: Trisyllabic production task

In the trisyllabic production task, participants were asked to read out a given list of existing trisyllabic stimulus words and phrases under three different conditions ("unswallowed", "swallowed", and "rhotacized"). Twenty-four stimulus sequences were employed for this task, with three extra sequences used for training. A complete list of the stimuli is provided in *Appendix 2*.

All stimuli were existing high-frequency trisyllabic sequences in Beijing Mandarin with a "medium-weak-strong" prosodic structure, the weak middle syllable of which contains a retroflex obstruent onset. The stimuli were controlled for the segments directly adjacent to the swallowing-triggering retroflex in the underlying form. That is, the nuclei of the two non-final syllables were monophthonic. The retroflex was preceded by either /a/, /i/ or /u/, and followed by either /ɚ/ or /u/. This forms six possible phonetic contexts surrounding the retroflex, evenly distributed across the twenty-four stimuli, namely, four stimuli for each phonetic context. These four stimuli that share the same phonetic context (i.e., the vowels adjacent to the swallowing-triggering retroflex) were further divided into two groups, forming two pairs that contrast in the manner of articulation (hereafter *MoA*) of the swallowing-triggering retroflex (fricative vs. affricate). This is to minimize the potential effect of MoA on the measurements. Since constructing trisyllabic minimal pairs that differ solely in the MoA of the retroflexes is difficult, only the onset on the initial syllable is controlled for, as it is directly adjacent to the rime of the initial syllable, which is the focus of the current study. A stimulus pair employed in the trisyllabic production task is provided in *Table 3*.

Table 3. Examples of stimuli employed in the trisyllabic production task

Stimulus	**Gloss**	**Transcription**	**MoA of retroflex**	**Vowel before retroflex**	**Vowel after retroflex**
八十九	'Eighty-nine'.	|ba^{55}. ʂɚ^{35}. tɕjoʊ^{214}|	fricative	/a/	/ɚ/
八只狗	'Eight dogs'	|ba^{55}. t͡ʂɚ^{55}. koʊ^{214}|	affricate	/a/	/ɚ/

3.2.2. Part 2: Sentence production task

In the sentence production task, participants were asked to read out a list of given sentences under different speech rates. Thirty stimulus sentences were employed for this task, with three additional sentences used for training. A complete list of the stimuli is provided in *Appendix 2*. All stimulus sentences contained a familiar trisyllabic location name in Beijing that has a retroflex onset in its middle syllable, in order to elicit more natural-speech like production. Unlike the trisyllabic production task, the location names were not controlled for the phonetic context surrounding the retroflex was not controlled, as it was hard to find enough location names with swallowing-triggering retroflex between monophthongs. The stimulus sentences followed the uniform structure in (9). An example of a stimulus sentence is provided in (10).

(9) Structure of stimulus sentences:

[Disyllabic person name] 在 [trisyllabic location name] 那边。

'[Disyllabic person name]'s house is nearby [trisyllabic location name].'

(10) Example of stimulus sentence:

李红家在西什库那边。

'Lihong's house is nearby Xishiku.'

3.3. Experimental design

3.3.1. Part 1: Trisyllabic production task

During the trisyllabic production task, participants produced each stimulus under three different conditions, namely, "unswallowed", "swallowed", and "rhotacized". The former two conditions were elicited by asking the participants to produce the sequence with or without "swallowing". The

“rhotacized” form here refers to the form in which the underlying trisyllabic sequence gets rhotacized in the rime of its initial syllable, and its middle syllable gets fully elided. This is to test the accuracy of this “swallowing” pattern of retroflex obstruents reported in previous studies.

Instructions on the condition in which participants were asked to produce the utterance were given in text, as summarized in *Table 4*, along with examples of the hypothetical surface forms under the three conditions of the underlying form |ba^{55}. ʂɚ^{35}. tɕjoʊ^{214}|. To note, as their phonetic nature remains to be tested, the “swallowed” and “rhotacized” surface forms in this table are only approximate transcriptions informed by previous studies.

Table 4. Instructions used to inform different experimental conditions in trisyllabic production task

Condition	**Instruction**		**Underlying form**	**Hypothetical surface forms**
Unswallowed	请您用正常语速大声朗读下方显示的三字词语，请注意： ‘Please read out loud the trisyllabic sequence displayed below at your normal speech rate. Please note that:’	不要吞音 ‘do not swallow’	|ba^{55}. ʂɚ^{35}. tɕjoʊ^{214}|	[ba^{55}. ʂɚ^{35}. tɕjoʊ^{214}]
Swallowed		请吞音 ‘please swallow’		[baɭ^{55}. tɕjoʊ^{214}]
Rhotacized		中间的音为儿化音 ‘there’s rhotacization in the middle’		[baɭ^{55}. tɕjoʊ^{214}]

Stimuli and instructions were presented with *PsychoPy* (Peirce et al., 2019). The *PsychoPy* interface presented to the participants for this task is as illustrated in *Figure 5*. Participants were instructed to produce the given trisyllabic sequence in the middle according to the instructions provided on the top of the screen after carefully reading the condition (see *Table 4*) written in red text. The experimenter clicked the “continue” button at the bottom of the screen once the participant confirmed to proceed to the next stimulus.

Figure 5. PsychoPy interface for the trisyllabic production task, English translation added for reference

3.3.2. Part 2: Sentence production task

During the sentence production task, participants produced each stimulus sentence at three different speech rates, namely, “normal”, “slow”, and “very slow”, plus an “unswallowed” condition to elicit production explicitly without “swallowing” as control. Instructions on the condition in which participants were asked to produce the utterance were given in text, summarized in *Table 5*. The *PsychoPy* interface for stimuli and instructions presentation is as illustrated in *Figure 6*. The most crucial difference of this task from the trisyllabic production task lied in the instructions given to the participants. That is, they were asked to imagine a casual, natural set-up under a Beijing Mandarin context when reading the stimulus sentences, for instance, talking with someone with a heavy Beijing accent. This is to elicit more natural-like speech and minimize the potential influence of Standard Mandarin on participants’ production.

Table 5. Instructions used to inform different experimental conditions in trisyllabic production task

Condition	**Instruction**	
Unswallowed	请您大声朗读下面的句子，语速： ‘Please read out loud the following sentence. The speech rate should be:’	正常 ‘normal’
Swallowed		较慢 ‘slow’

Rhotacized		很慢 ‘very slow’
Unswallowed	请您用正常语速大声朗读下方显示的句子，请注意： ‘Please read out loud the sentence displayed below at your normal speech rate. Please note that:’	不要吞音 ‘do not swallow’

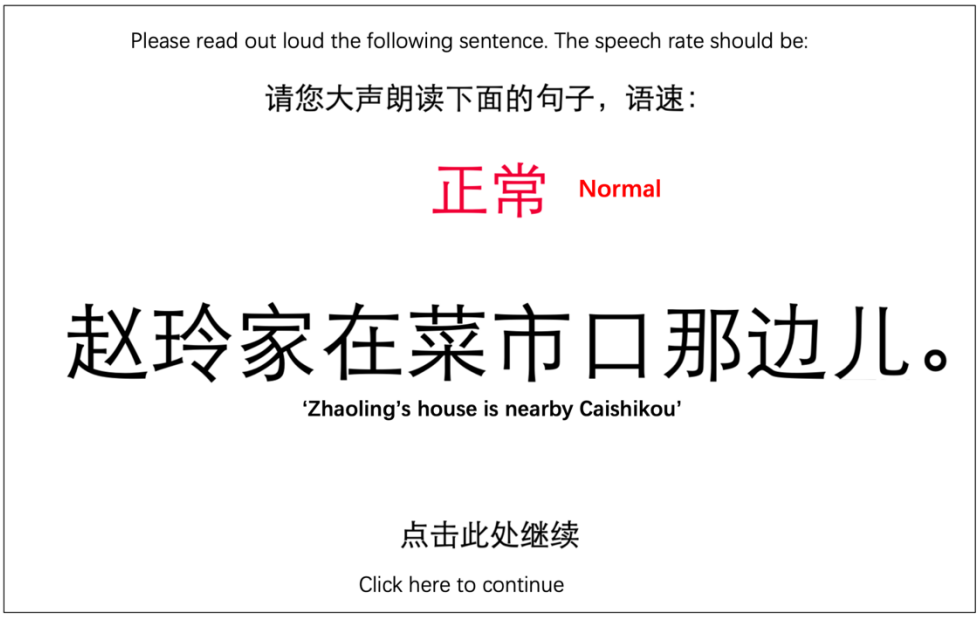

Figure 6. *PsychoPy* interface for the sentence production task, English translation added for reference

3.4. Procedure

Data collection was conducted in person in Beijing. Prior to taking part in the experiment, participants were briefly informed of the purpose of the study and provided their informed consent to participate and be recorded. The information brochure and consent form are provided in *Appendix 4*.

During the experiment, participants were seated in front of a desk in a quiet room, with a MacBook laptop on the desk for displaying stimulus sequences or sentences as well as corresponding instructions. Recording was made with an *EDIROL R-1 Portable WAVE/MP3* recorder and a *Beyerdynamic TG H34c* microphone headset.

The entire procedure with each participant lasted approximately 40 minutes. Participants first took the trisyllabic production task. To familiarize participants with the task, they were first trained with three sample stimuli, each presented in the three experimental conditions in the order of "unswallowed", "swallowed" and "rhotacized" (i.e., nine trail utterances in total). Recording started after the completion of the training. For the recorded part, stimuli were presented in randomized order to avoid priming the participants with any of the conditions or stimuli. The presented order was automatically logged in a .csv file to facilitate data processing. The procedure of the subsequent sentence production task was largely consistent with the trisyllabic production task, only substituting trisyllabic sequences with sentences. Similarly, participants were trained with three sample sentences before the recording, each presented in the three speech rates in the order of "normal", "slow" and "very slow".

3.5. Data processing & measurements

3.5.1. Pre-processing

The raw audio recordings were first manually edited in *Praat* (Boersma & Weenink, 2025) to remove invalid parts, such as participants talking to the experimenter, stutters, or instances where participants misread the instructions and were asked to produce the utterance again. Other extraneous noises visible in the waveform and the spectrogram, such as heavy breathing or coughing, were also cut. Only the target utterances were retained in the cleaned audio files. The sounding parts of each processed recording were then marked out in an .TextGrid file with the "Trim silences…" function in Praat, in preparation for annotation.

3.5.2. Data processing & measurements

3.5.2.1. Trisyllabic production task

Preliminary annotation of the audio files collected in the trisyllabic production task was conducted using a *Praat* script, which automatically inserted the text of the stimulus sequences and corresponding conditions logged in the .csv files into two newly created interval tiers in the .TextGrid file obtained from pre-processing, named "Word" and "Condition". Two other

interval tiers named “Non-final” and “Onset/rime” were also added. Detailed annotation was conducted fully manually. The *Montreal Forced Aligner* was tested for semi-automatic annotation (McAuliffe et al., 2017). However, it was not fully applicable as the tool lacks an acoustic model and dictionary for “Beijing Swallowing”. For each trisyllabic utterance, an interval starting with the onset of its initial syllable and ending with the onset of its final syllable was annotated on the “Non-final” tier, labeled “nf”. This marks the non-final-syllable portion (referred to as *non-final part* hereafter) of each utterance. The bold and underlining in *Table 6* illustrates the non-final part in the surface form under the three experimental conditions. Another interval starting with the onset of the rime in the initial syllable and ending with the offset of the rime in the initial syllable of each utterance was annotated on the “Onset/rime” tier, labeled “r1”.

Table 6. Non-final part in the surface form under three experimental conditions, marked with bold and underline

Underlying form	**Condition**	**Hypothetical surface form**
|ba^{55}. ʂə^{35}. tɕjoʊ^{214}|	Unswallowed	[**ba^{55}. ʂə^{35}.** tɕjoʊ^{214}]
	Swallowed	[**baɻ^{55}.** tɕjoʊ^{214}]
	Rhotacized	[**baɻ^{55}.** tɕjoʊ^{214}]

Data extraction for the trisyllabic production task was carried out using a *Praat* script. The following measurements were included:

1) Number of syllable(s)[4]: The number of syllables in the non-final part of each utterance was measured. This was achieved through the “To PointProcess…” function in *Praat* by detecting long intervals between adjacent glottal pulses. To achieve more precise glottal pulses, the pitch floor was set at 100 Hz for female speakers and 75 Hz for male speakers, while the pitch ceiling was 500 Hz for the female group and 350 Hz for the male group. Specifically, if the duration of an interval exceeded 0.05 seconds and occurred within the non-final part of the utterance, it was considered to indicate the presence of a syllable boundary, as all investigated segments were voiceless and there was not intervocalic voicing.

[4] This method was developed based on speech data from the trisyllabic production task of Participant 01.

In other words, the non-final part was deemed monosyllabic if no such long interval was detected, otherwise, the non-final part was considered disyllabic.

2) Durations: The duration of the non-final part of each utterance was measured.

3) F3 values: The approximate trajectory of the F3 in the initial syllable's rime of each utterance was measured with the "To Formant (burg)" function in *Praat*. In order to achieve more precise formant measurement, the formant ceiling was set at 5500 Hz for female speakers, and 5000 Hz for male speakers. To capture the F3 curve, the duration of the initial syllable's rime was divided into five equal intervals (0–20%, 20–40%, 40–60%, 60–80%, 80–100%), and the average F3 values were extracted for each interval. The overall average F3 value of the initial syllable's rime of each utterance was also extracted.

3.5.2.2. Sentence production task

Data processing for the sentence production task was also conducted through a combination of *Praat* scripts and manual annotation. The procedures, as well as corresponding *Praat* settings, were largely identical to the trisyllabic production task (see *Section 3.5.2.1*). The non-final part of the targeted trisyllabic location name in each utterance was manually marked out, measured for its duration and the number of syllables it contains. F3 measurement was not conducted for this task as the targeted location names were not controlled for the phonetic context of the swallowing-triggering retroflex.

4. Results

Data analysis and visualization were conducted in *R* (R Core Team, 2025), with the *ggplot2* package (Wickham, 2016) for visualization, and the *lmerTest* package (Kuznetsova et al., 2017) for regression analyses. Outliers were removed according to the 1.5 interquartile range (IQR) rule. Namely, values more than 1.5 IQR below the first quartile (Q1) or above the third quartile (Q3) were excluded from the dataset (Maini, 2025).

4.1. Part1: Trisyllabic production task

4.1.1. Number of syllable(s) in non-final part of recorded utterances

Figure 7 presents a stacked bar plot illustrating the proportion of monosyllables and disyllables in the non-final part of the recorded utterances under the "unswallowed" and "swallowed" conditions. Corresponding descriptive statistics are provided in *Table 7*. The proportion of monosyllables was higher under the "swallowed" condition (93%) than the "unswallowed" condition (13%).

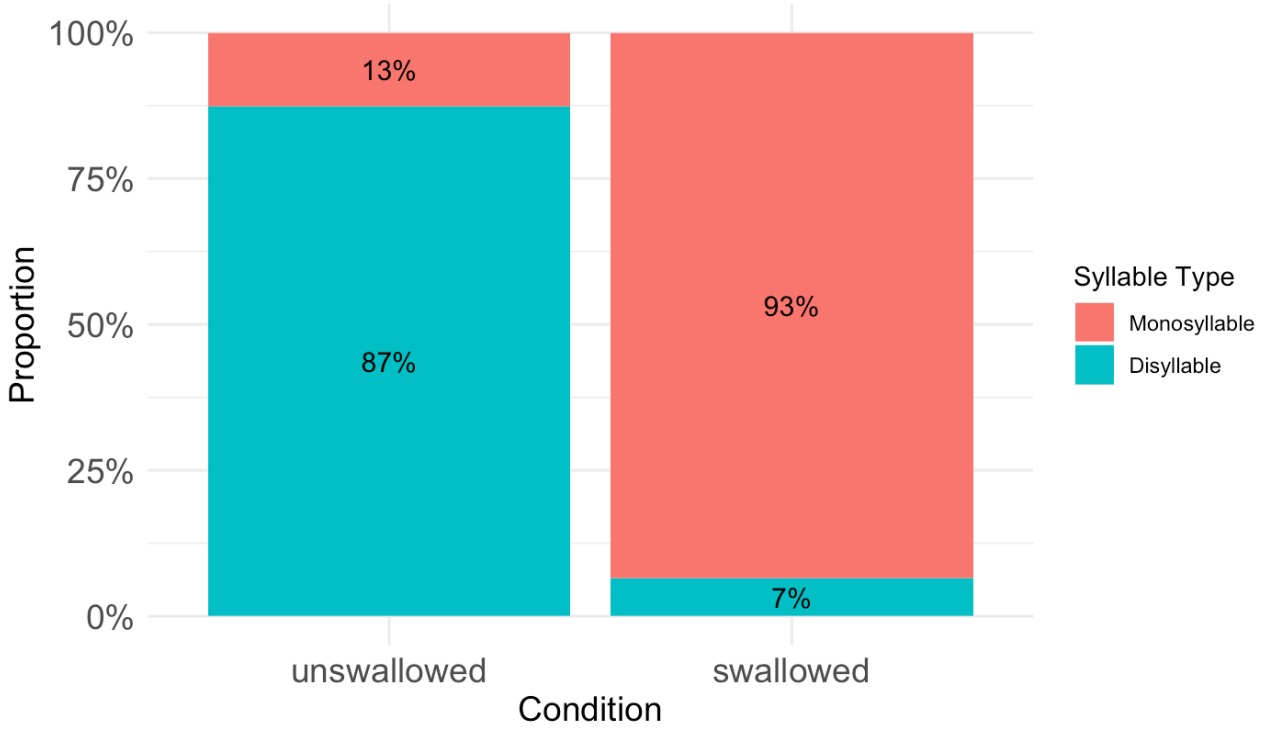

Figure 7. Stacked bar plot showing the proportion of monosyllables and disyllables in the non-final part under the "unswallowed" and "swallowed" conditions.

Table 7. Descriptive statistics of monosyllables and disyllables proportions in the non-final part under "unswallowed" and "swallowed" conditions.

Condition	**n (total)**	**n (monosyllable)**	**n (disyllable)**	**Prop. (monosyllable)**	**Prop. (disyllable)**
Unswallowed	355	45	310	13%	87%
Swallowed	352	329	23	93%	7%

A generalized linear mixed-effects model was fitted to analyze the effect of condition

("unswallowed" vs. "swallowed") on the proportion of monosyllables and disyllables in the non-final part using the binomial family with a logit link function, with condition as the fixed effect. A contrast coding scheme was applied to the condition variable, coded as "unswallowed = –1/2, swallowed = +1/2". Random intercepts were specified for both participants and tokens to account for individual- and stimulus-specific variability. The *R* code for the model is provided in (11). The results, as summarized in *Table 8*, showed that the proportion of monosyllables is significantly higher under the "swallowed" condition compared to the "unswallowed" condition (Estimate= -5.45, SE = 0.38, z = -14.16, p < .001).

(11) model <- glmer (num.syllables ~ condition + (1 | participant) + (1 | token), data = df, family = binomial)[5]

Table 8. Summary of fixed effects in the generalized linear mixed-effects model on number of syllables in the non-final portion of the trisyllabic sequences, grouped by condition ("unswallowed" vs. "swallowed"). Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Factor	**Estimate**	**Std. Error**	**z-value**	**p-value**	**Sig.**
(Intercept)	-0.39	0.31	-1.28	0.209	
condition+Uns-Swa	-5.45	0.38	-14.16	<2e-16	***

4.1.2. Durations of recorded utterances' non-final part

Figure 8 presents violin plots with overlaid boxplots illustrating the distribution of the duration of the non-final part of the recorded utterances under three experimental conditions (i.e., "unswallowed", "swallowed", and "rhotacized"). Outliers were removed according to the 1.5 IQR rule per condition. Corresponding descriptive statistics are provided in *Table 9*. The average duration is longest under the "unswallowed" condition (M = 0.405, SD = 0.084), followed by "swallowed" (M = 0.236, SD = 0.076) and "rhotacized" (M = 0.234, SD = 0.081).

[5] To note, the model code presented in the results section is modified for the name of the data frame (renamed as "df"). The rest of the code is identical to that in the original script.

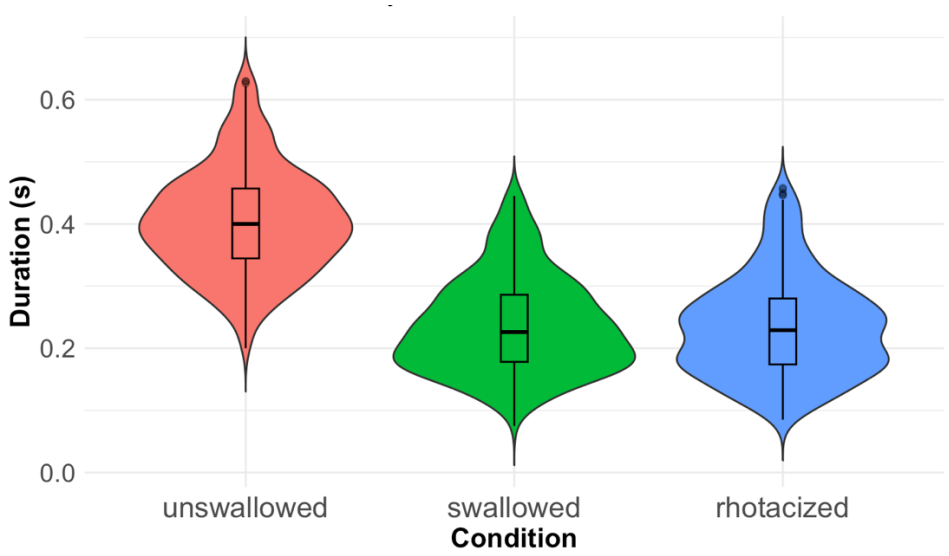

Figure 8. Violin plot with overlaid boxplot showing the distribution of non-final part durations across conditions

Table 9. Descriptive statistics of non-final part durations (in seconds) by condition

Condition	**Mean**	**SD**	**Min**	**Q1**	**Median**	**Q3**	**Max**	**n**
Unswallowed	0.405	0.084	0.200	0.345	0.400	0.457	0.630	351
Swallowed	0.236	0.076	0.075	0.178	0.226	0.286	0.445	357
Rhotacized	0.234	0.081	0.085	0.174	0.229	0.280	0.458	353

A linear mixed-effects model was fitted to analyze the effect of condition on the non-final part duration, with condition as the fixed effect. Forward difference coding was applied for the 3-level variable *condition* with the contrasts shown in *Table 10*, while the binary factor *MoA* was coded as "fricative = –1/2, affricate = +1/2". Random intercepts and slopes for *condition* and *MoA* were specified for both participants and tokens to account for individual- and stimulus-specific variability. The *R* code for the model is provided in (12).

(12) model <- lmer (dur.non.fin ~ condition + (1 + condition + MoA | participant) + (1 + condition + MoA | token), data = df)

Table 10. Forward difference coding scheme for the 3-level variable *condition*

	+Uns-Sw	+Sw-Rho
Unswallowed	2/3	1/3
Swallowed	-1/3	1/3
Rhotacized	-1/3	-2/3

The results, as summarized in *Table 11*, show that the "swallowed" condition has significantly shorter duration compared to the "unswallowed" condition (Estimate= 0.17, SE = 0.01, t = 14.44, p < .001), while the difference between the "swallowed" condition and the "rhotacized" condition is not statistically significant (Estimate = -0.00, SE = 0.01, t = -0.01, p > .1).

Table 11. Summary of fixed effects in the linear mixed-effects model on non-final part duration. Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Factor	**Estimate**	**SE**	**t-value**	**p-value**	**Sig.**
(Intercept)	0.30	0.02	18.95	1.26e-14	***
condition+Uns-Sw	0.17	0.01	14.44	1.47e-10	***
condition+Sw-Rho	-0.00	0.01	-0.01	0.989	

4.1.3. F3 values

1) F3 trajectories

The F3 trajectories in the initial syllable's rime in the recorded utterances under the three experimental conditions are illustrated in *Figure 9*, *Figure 10*, and *Figure 11*, grouped by gender and the phonetic context of the swallowing-triggering retroflex (to note, the "rhotacized" condition is only grouped by the vowel preceding the retroflex, instead of the combination of preceding and following vowels, as the underlyingly middle syllable is absent in the surface form under this condition). Outliers were also removed according to the 1.5 IQR rule per combination of condition, context, gender and interval. Given the scope of the present study, these trajectories are included only for illustrative purposes and to inform other analyses but were not examined in detail.

a) Vowel before swallowing-triggering retroflex: /a/

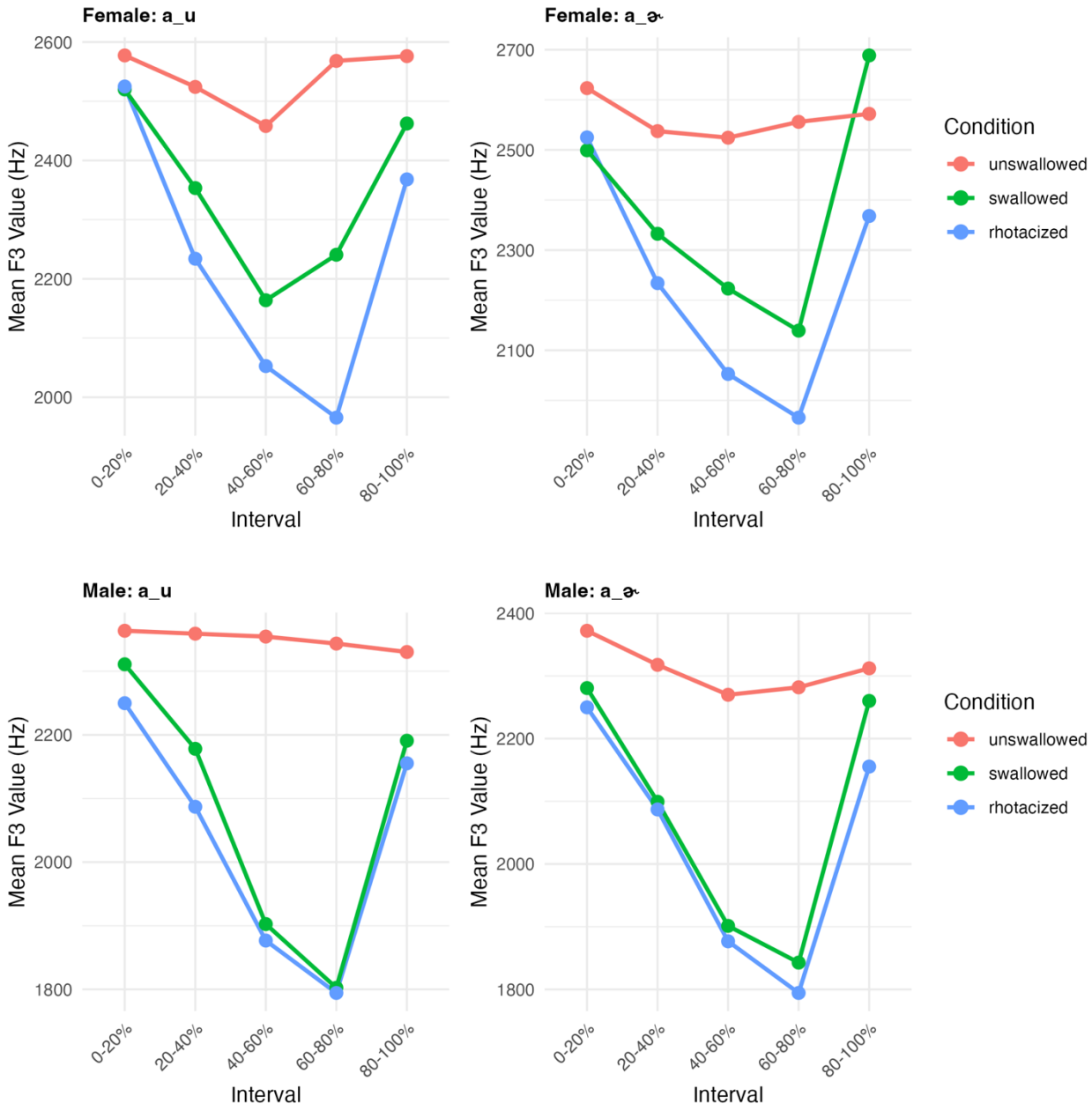

Figure 9. Averaged approximate F3 trajectories in the initial syllable's rime, the swallowing-triggering retroflex segment preceded by /a/, grouped by gender and the following vowel

b) Vowel before swallowing-triggering retroflex: /i/

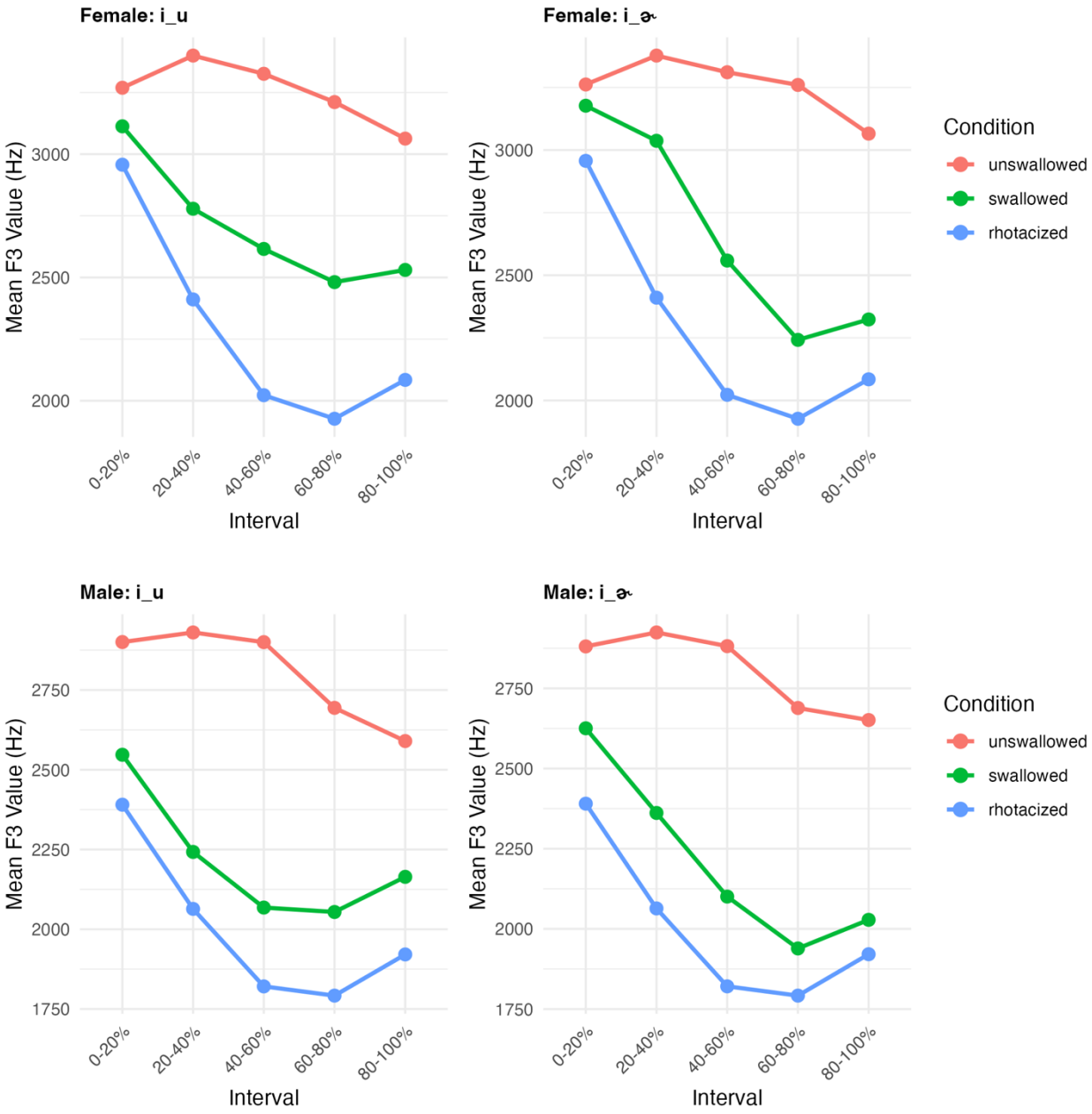

Figure 10. Averaged approximate F3 trajectories in the initial syllable's rime, the swallowing-triggering retroflex segment preceded by /i/, grouped by gender and the following vowel

c) Vowel before swallowing-triggering retroflex: /u/

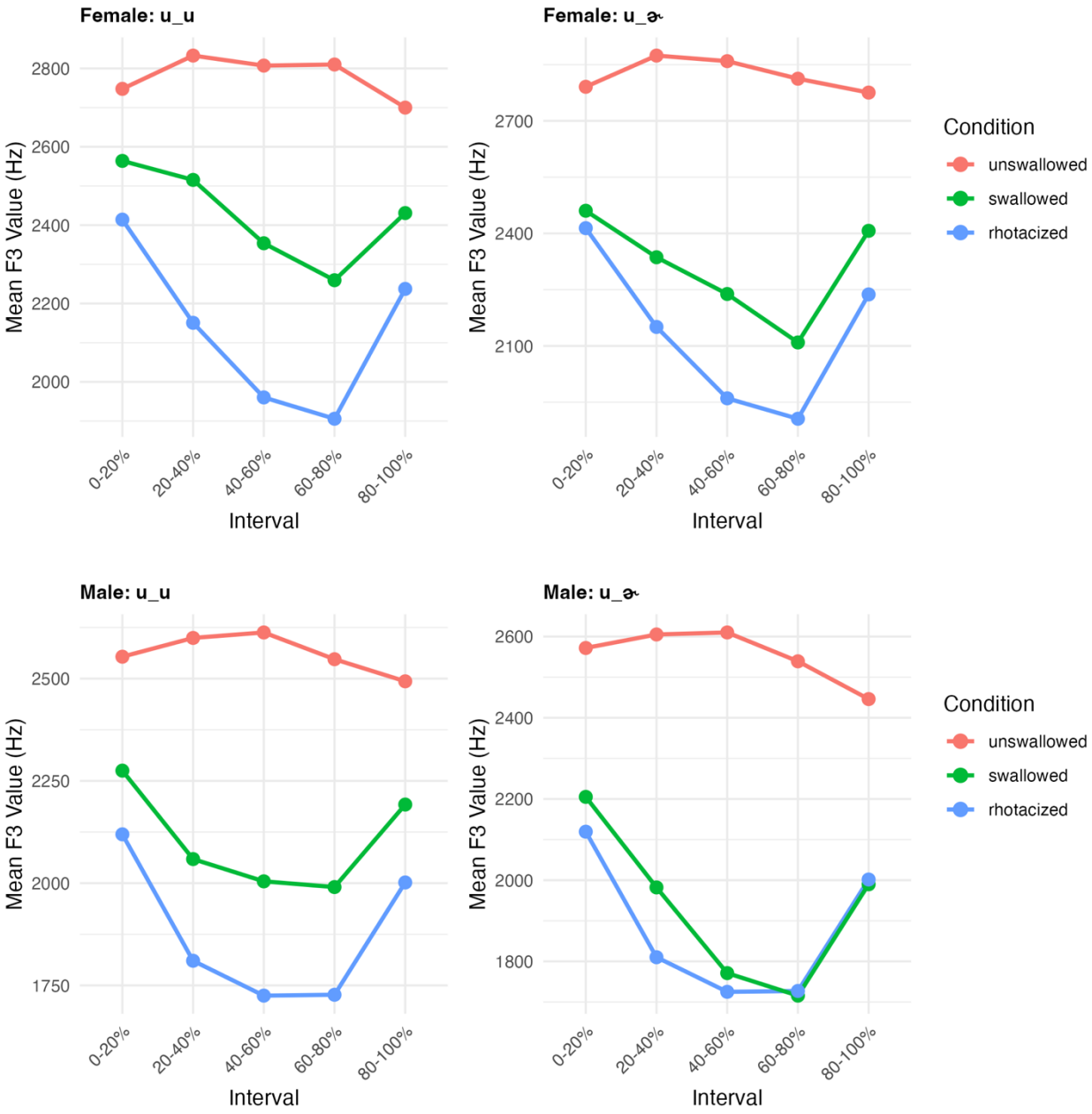

Figure 11. Averaged approximate F3 trajectories in the initial syllable's rime, the swallowing-triggering retroflex segment preceded by /u/, grouped by gender and the following vowel

2) Average F3 values

Since direct comparison of the formant trajectories is beyond the scope of the present analysis, the F3 values in the initial syllable's rime of the recorded utterances were analyzed using their mean to examine the overall effect of "swallowing" and "rhotacization" on F3 value.

Figure 12 presents violin plots with overlaid boxplots illustrating the distribution of the average F3 values of initial syllable's rime across the three experimental conditions, grouped by gender and phonetic context. Outliers were also removed according to the 1.5 IQR rule per combination of gender, condition and context (the "rhotacized" condition was only grouped by the vowel preceding the retroflex). The corresponding descriptive statistics, in the interest of space, are provided in *Appendix 3*. In general, it can be observed that the "unswallowed" condition has the highest mean F3 values, followed by "swallowed" and "rhotacized".

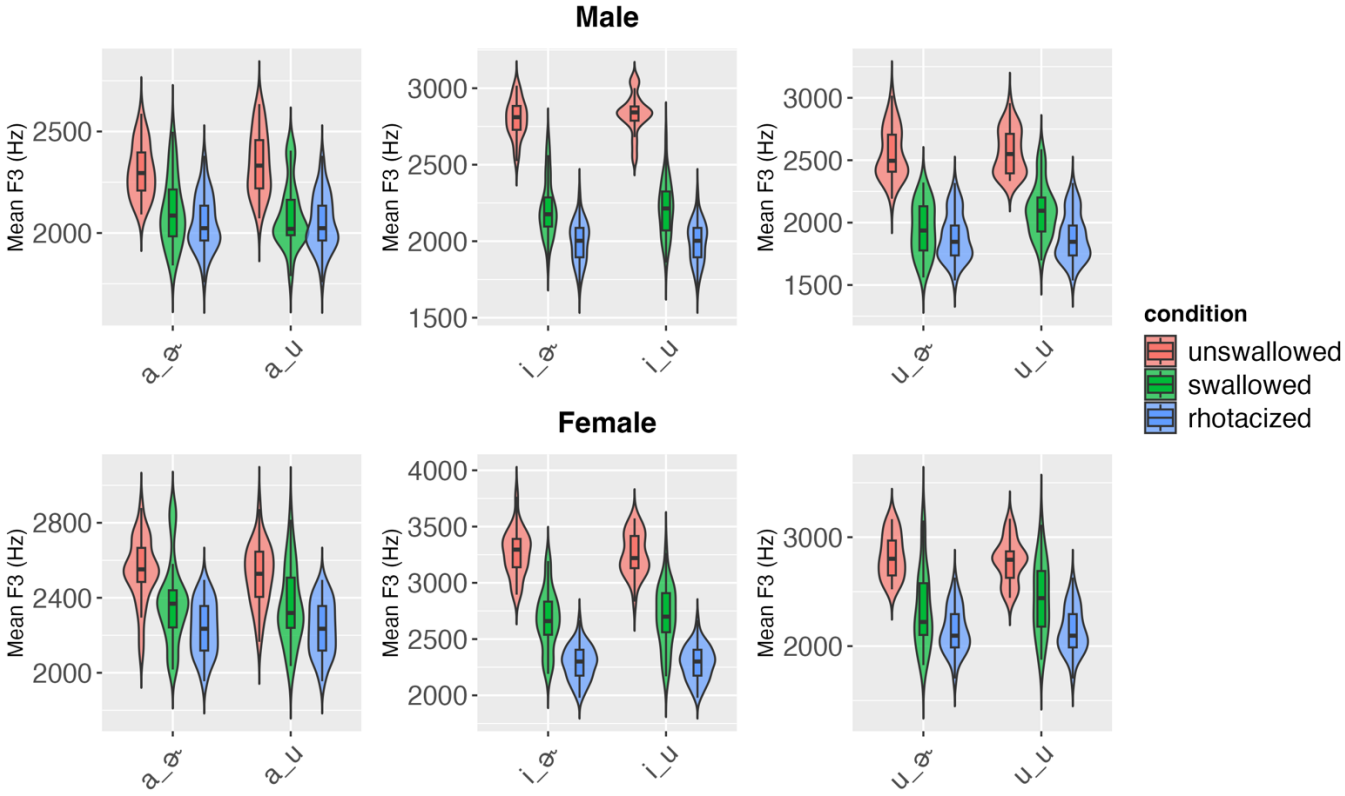

Figure 12. Violin plot with overlaid boxplot showing the distribution of average F3 values of initial syllable's rime across the three conditions, grouped by gender and phonetic context

A linear mixed-effects model was fitted to analyze the effect of condition on the average F3 value of initial syllable's rime, with condition as the fixed effect. Forward difference coding applied for the 3-level variable *condition* was identical to that illustrated in *Table 10*. Random intercepts and slopes for condition were specified for both participants and tokens to account for individual- and stimulus-specific variability. The *R* code for the model is provided in (13). The models were fitted separately for each combination of gender and context.

(13) model <- lmer(F3 ~ condition + (1 + condition| participant) + (1 + condition| token), data = df)

The results, as summarized in *Table 12*, show that the "unswallowed" condition has significantly higher average F3 values compared to the "swallowed" and "rhotacized" conditions across both genders and all contexts. The average F3 values are also generally significantly higher under the "swallowed" condition than "rhotacized". The difference between these two conditions is not statistically significant only under the "u_ɚ" context for the female group (Estimate= 134.04, SE = 91.37, t = 1.47, p > .1), and under "a_u" (Estimate= 44.79, SE = 53.89, t = 0.83, p > .1) and "u_ɚ" (Estimate= 54.81, SE = 47.49, t = 1.15, p > .1) contexts for the male group. These results are consistent with the general patterns observed in the averaged F3 trajectories. (see *Figure 9*, *Figure 10*, and *Figure 11*).

Table 12. Summary of fixed effects in the linear mixed-effects model on average F3 value in initial syllable rimes, grouped by gender and context. Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Gender	Context	Condition	Estimate	SE	t-value	p-value	Sig.
Female	a_u	Intercept	2367.34	31.65	74.79	7.41e-11	***
		Uns-Sw	169.68	68.57	2.47	0.045	*
		Sw-Rho	112.52	51.78	2.17	0.067	.
	a_ɚ	Intercept	2388.31	37.30	64.02	7.79e-12	***
		Uns-Sw	164.21	40.79	4.03	0.000	***
		Sw-Rho	146.71	50.89	2.88	0.021	*
	i_u	Intercept	2751.53	43.72	62.94	1.3e-11	***
		Uns-Sw	545.56	118.34	4.61	0.002	**
		Sw-Rho	423.05	93.69	4.52	0.004	**
	i_ɚ	Intercept	2749.96	42.94	64.05	3.07e-11	***

		Uns-Sw	592.22	120.95	4.90	0.002	**
		Sw-Rho	397.91	78.13	5.09	0.001	***
	u_u	Intercept	2447.02	48.71	50.24	1.67e-09	***
		Uns-Sw	346.33	80.02	4.33	0.007	**
		Sw-Rho	282.84	86.66	3.26	0.011	*
	u_ɚ	Intercept	2411.81	65.21	36.99	8.78e-12	***
		Uns-Sw	538.30	108.75	4.95	0.002	**
		Sw-Rho	134.04	91.37	1.47	0.188	
Male	a_u	Intercept	2162.85	41.77	51.78	9.9e-11	***
		Uns-Sw	250.43	45.77	5.47	0.00473	.
		Sw-Rho	44.79	53.89	0.83	0.442	
	a_ɚ	Intercept	2156.44	36.99	58.30	6.86e-13	***
		Uns-Sw	192.23	45.80	4.20	0.004	**
		Sw-Rho	64.27	28.85	2.23	0.052	.
	i_u	Intercept	2347.64	34.46	68.12	1.41e-11	***
		Uns-Sw	594.73	51.77	11.49	0.000	***
		Sw-Rho	230.95	46.24	5.00	0.003	**
	i_ɚ	Intercept	2334.30	40.53	57.60	4.23e-13	***
		Uns-Sw	602.17	51.60	11.67	2.60e-06	***
		Sw-Rho	206.45	44.91	4.60	0.002	**
	u_u	Intercept	2181.31	44.29	49.25	2.55e-10	***
		Uns-Sw	471.72	83.51	5.65	0.001	***
		Sw-Rho	209.92	80.48	2.61	0.044	*
	u_ɚ	Intercept	2126.18	43.52	48.85	1.70e-14	***
		Uns-Sw	615.17	91.77	6.70	3.83e-05	***
		Sw-Rho	54.81	47.49	1.15	0.278	

3) Effect of the vowel following the swallowing-triggering retroflex

This section examines the average F3 values in the initial syllable's rime under the "swallowed" condition, grouped by gender and the vowel following the swallowing-triggering retroflex. This is to examine whether the vowel following the retroflex is fully elided during "swallowing" or leaves acoustic traces on F3 value.

Figure 13 presents violin plots with overlaid boxplots illustrating the distribution of the average F3 values in the initial syllable's rime under the "swallowed" condition, grouped by gender and the following vowel. Outliers were removed according to the 1.5 IQR rule per combination of gender

and following vowel. Corresponding descriptive statistics are provided in *Table 13*. The average F3 is higher when the retroflex is followed by /u/ than followed by /ɚ/ for both genders. Specifically, for the female group, the mean average F3 value is 2476.61 Hz when followed by /ɚ/ (SD = 333.26), while being 2520.60 Hz when followed by /u/ (SD = 327.95). For the male group, the mean average F3 value is 2077.54 Hz when followed by /ɚ/ (SD = 216.15), while being 2142.65 Hz when followed by /u/ (SD = 202.83).

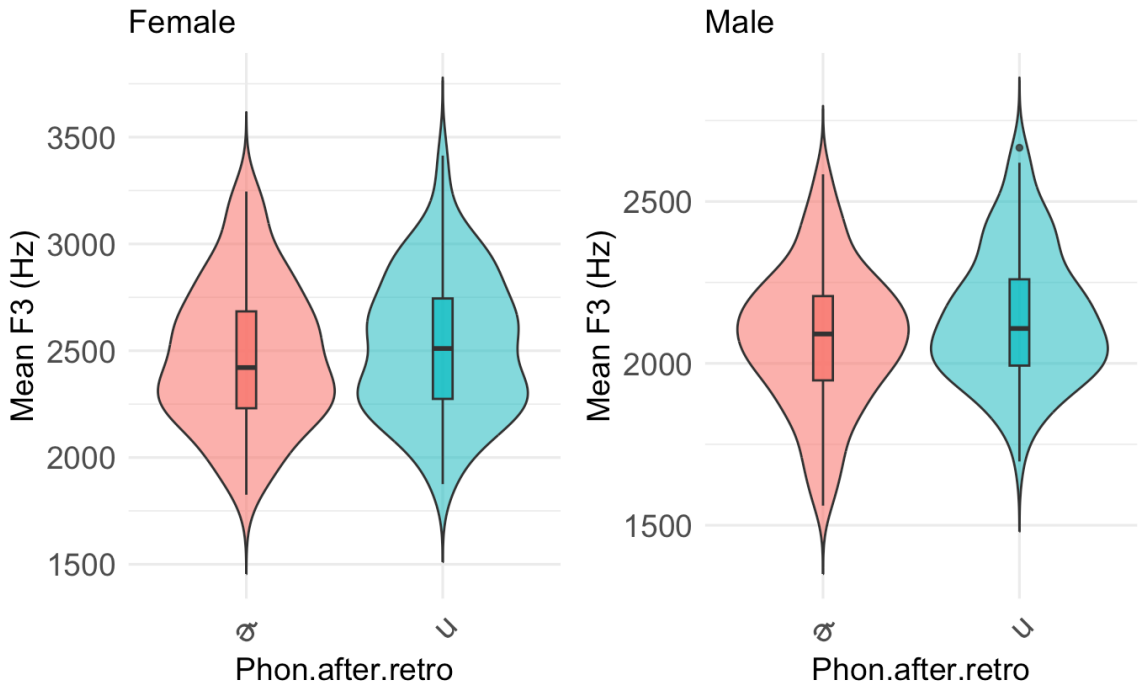

Figure 13. Violin plot with overlaid boxplot showing the distribution of average F3 values of initial syllable rime under the "swallowed" condition, grouped by gender and the vowel following the swallowing-triggering retroflex segment.

Table 13. Descriptive statistics of average F3 values of initial syllable rime (in Hz), grouped by gender and following vowel

Gender	Following vowel	Mean	SD	Min	Q1	Median	Q3	Max	n
Female	/ɚ/	2476.61	333.26	1825.70	2230.62	2421.08	2684.23	3245.60	84
	/u/	2520.60	327.95	1875.51	2274.48	2510.51	2744.81	3413.59	84
Male	/ɚ/	2077.54	216.15	1560.59	1947.52	2090.75	2207.85	2583.99	95
	/u/	2142.65	202.83	1697.03	1993.15	2107.83	2259.71	2666.05	94

A linear mixed-effects model was fitted to analyze the effect of following vowel (/ɚ/ vs. /u/) on average F3 values in the initial syllable's rime under the "swallowed" condition. The fixed effect was the following vowel, which was contrast coded as "ɚ = –1/2, u = +1/2". Random intercepts for both participants and tokens were specified. The *R* code for the model is provided in (14). The models were fitted separately for each gender. The results are summarized in *Table 14*. The effect of following vowel on F3 was not statistically significant for both genders (Female: Estimate = 43.98, SE = 77.63, t = 0.57, p = > .01; Male: Estimate = 67.67, SE = 47.26, t = 1.43, p > .01).

(14) model <- lmer (F3 ~ phon.after.retro + (1 | participant) + (1 | token), data = df)

Table 14. Summary of fixed effects in the linear mixed-effects model on average F3 value in initial syllable rimes under "swallowed" condition, grouped by gender and following vowel. Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Gender	**Factor**	**Estimate**	**SE**	**t-value**	**p-value**	**Sig.**
Female	(Intercept)	2498.61	58.32	42.846	5.76e-14	***
	/ɚ/ - /u/	43.98	77.63	0.567	0.577	
Male	(Intercept)	2112.76	50.45	41.880	1.19e-12	***
	/ɚ/ - /u/	67.67	47.26	1.432	0.166	

However, as shown in *Figure 9*, *Figure 10*, and *Figure 11*, the formant trajectories under the "swallowed" condition exhibit noticeable differences when followed by different vowels. Therefore, the effect of the following vowel on the average F3 value in the 60–80% time interval of the initial syllable's rime under the "swallowed" condition was further analyzed. The 60-80% interval was selected because the F3 trajectories show the clearest visual distinction between the two following vowels across almost all contexts for both genders in this portion.

Figure 14 presents violin plots with overlaid boxplots illustrating the distribution the average F3 values in the 60-80% time interval of the initial syllable's rime under the "swallowed" condition, grouped by gender, preceding vowel and following vowel. Outliers were removed according to the 1.5 IQR rule per combination of gender and context. In the interest of space, corresponding descriptive statistics are provided in *Appendix 3*. Overall, the average F3 value in this interval is

higher when followed by /u/ than by /ɚ/ for both genders across nearly all contexts, except when preceded by /a/ in the male group.

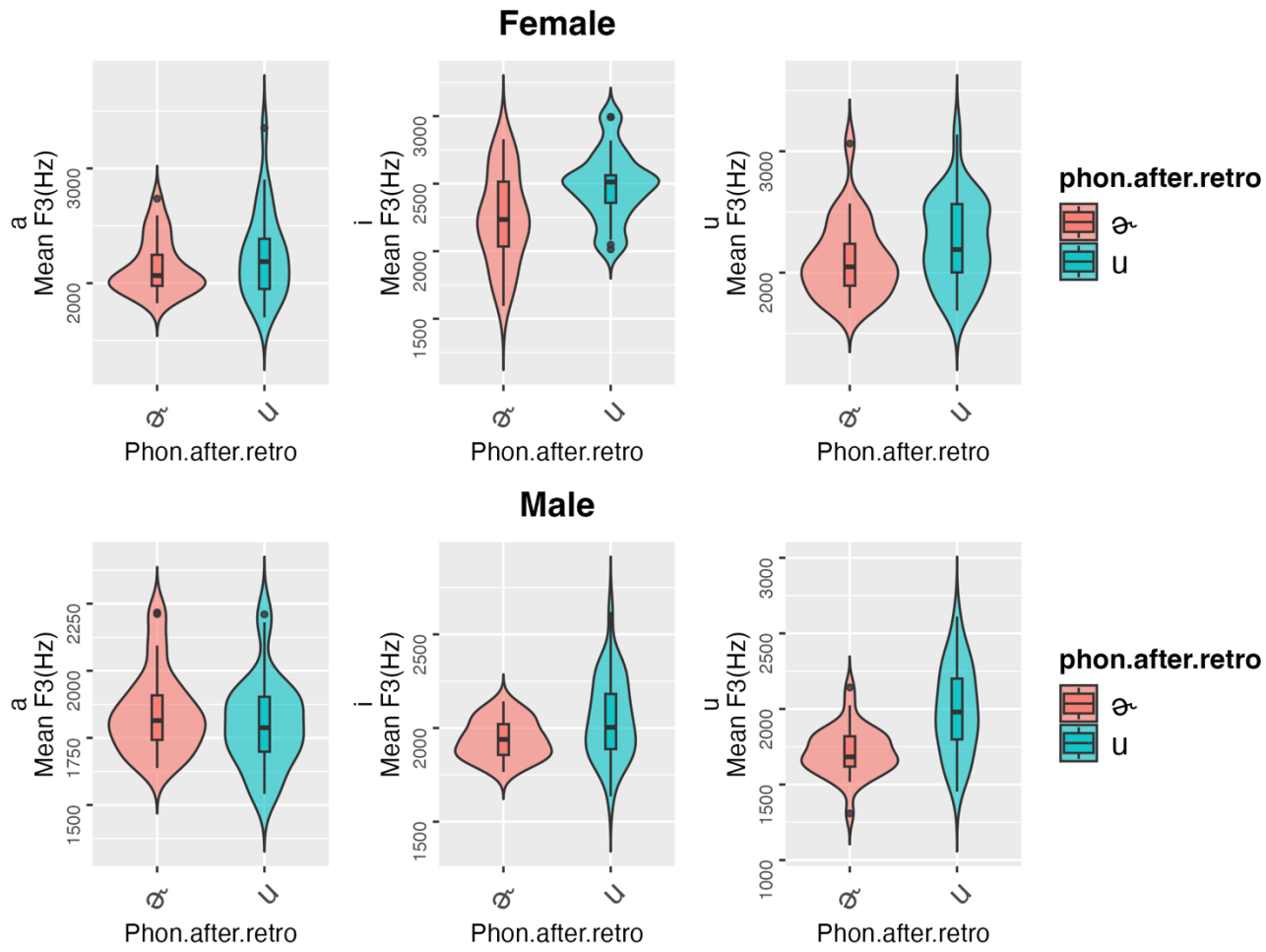

Figure 14. Violin plot with overlaid boxplot showing the distribution of average F3 values in the 60-80% time interval of initial syllable's rime under the "swallowed" condition, grouped by gender, preceding vowel and following vowel

A linear mixed-effects model was fitted to analyze the effect of following vowel (/ɚ/ vs. /u/) on average F3 values in the 60-80% time interval under the "swallowed" condition. The model and contrast coding were identical to those used for the overall average F3 analysis (see above). The models were fitted separately per combination of gender and preceding vowel. The results are summarized in *Table 15*. For female speakers, the following vowel /u/ significantly increased the F3 value (Estimate = 224.00, SE = 58.23, t = 3.85, $p < .001$) compared to /ɚ/ when the preceding vowel is /i/. For male speakers, the following vowel /u/ significantly increased the F3 value

compared to /ɚ/ when the preceding vowel is either /i/ (Estimate = 112.24, SE = 41.33, t = 2.72, p < .01) or /u/ (Estimate = 273.17, SE = 81.26, t = 3.36, p < .05).

Table 15. Summary of fixed effects in the linear mixed-effects model on average F3 value in initial syllable rimes under "swallowed" condition in the 60-80% time interval, grouped by gender, preceding vowel and following vowel. Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Gender	Preceding vowel	Factor	Estimate	SE	t-value	p-value	Sig.
Female	/a/	(Intercept)	2189.91	52.19	41.96	1.27e-06	***
		/ɚ/ - /u/	104.74	94.40	1.11	0.307	
	/i/	(Intercept)	2359.78	92.77	25.44	2.49e-07	***
		/ɚ/ - /u/	224.00	58.23	3.85	0.000	***
	/u/	(Intercept)	2188.49	75.77	28.88	4.22e-08	***
		/ɚ/ - /u/	134.63	104.05	1.29	0.242	
Male	/a/	(Intercept)	1822.64	21.49	84.83	<2e-16	***
		/ɚ/ - /u/	-40.30	42.97	-0.94	0.353	
	/i/	(Intercept)	1996.51	31.55	63.28	9.05e-11	***
		/ɚ/ - /u/	112.24	41.33	2.72	0.009	**
	/u/	(Intercept)	1853.86	53.53	34.63	3.71e-10	***
		/ɚ/ - /u/	273.17	81.26	3.36	0.015	*

4.2. Part 2: Sentence production task

4.2.1. Number of syllable(s) in the non-final part of the targeted location names

Figure 15 presents a stacked bar plot illustrating the proportion of monosyllables and disyllables in the non-final part of the targeted location names under the four conditions. Corresponding descriptive statistics are provided in *Table 16*. The proportion of monosyllables was lowest under the "unswallowed" condition (26%), followed by "very solw" (44%), "slow" (52%) and "normal" (73%).

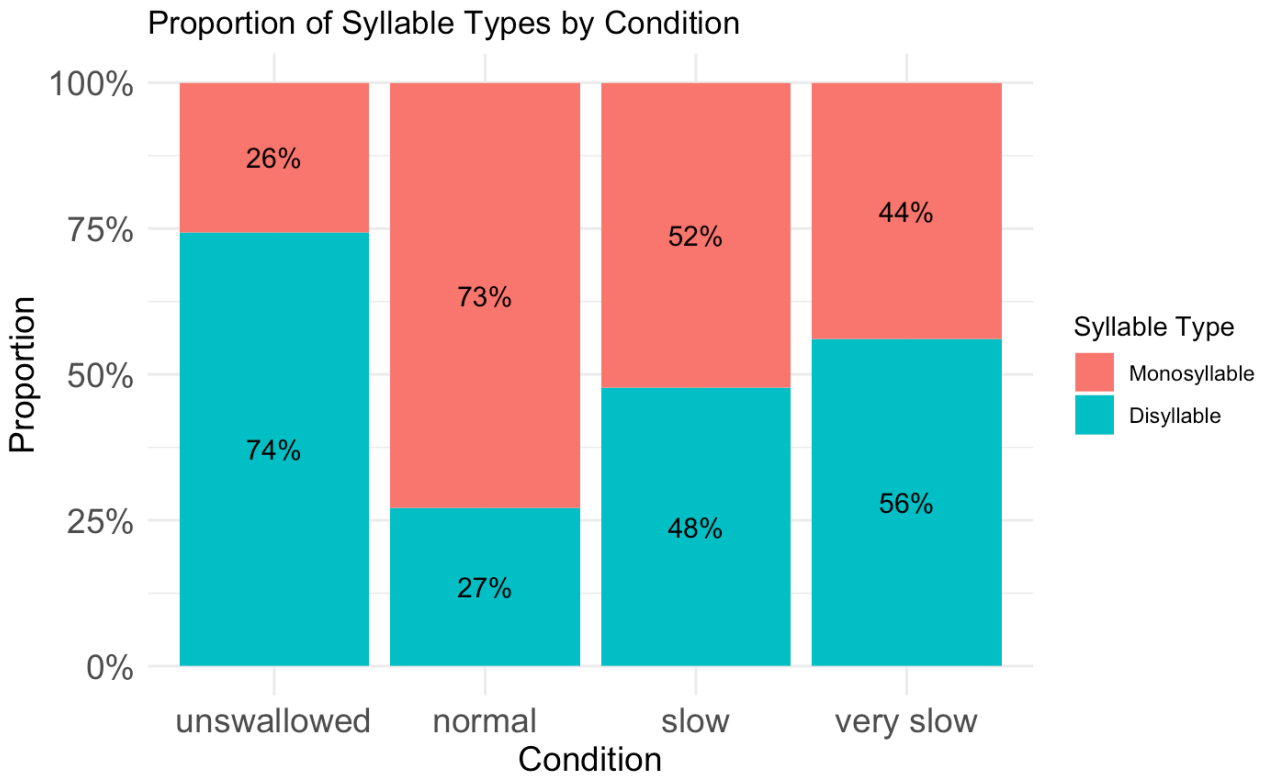

Figure 15. Stacked bar plot showing the proportion of monosyllables and disyllables in the non-final part of the targeted location names across the four experimental conditions

Table 16. Descriptive statistics of monosyllables and disyllables proportions in the non-final part of the targeted location names under the four conditions.

Condition	**n (total)**	**n (monosyllable)**	**n (disyllable)**	**Prop. (monosyllable)**	**Prop. (disyllable)**
Unswallowed	148	38	110	26%	74%
Normal	148	108	40	73%	27%
Slow	149	78	71	52%	48%
Very slow	150	66	84	44%	56%

A generalized linear mixed-effects model was fitted to analyze the effect of speech rate on the proportion of monosyllables and disyllables in the non-final part using the binomial family with a logit link function, with condition as the fixed effect. The default treatment coding scheme of the *lmertest* package was applied to the variable condition. Random intercepts were specified for both participants and tokens to account for individual- and stimulus-specific variability. The *R* code for the model is provided in (15). The results , as summarized in *Table 17*, showed that the proportion of monosyllables is significantly lower under the “unswallowed” condition compared to the

"normal" (Estimate= -2.62, SE = 0.31, -8.40, p < .001), "slow" (Estimate= -1.46, SE = 0.29, z = -5.13, p < .001) and "very slow" (Estimate= -1.04, SE = 0.28, 3.69, p < .001) conditions.

(15) model <- glmer (num.syllables ~ condition + (1 | participant) + (1| token), data = df, family = binomial)

Table 17. Summary of fixed effects in the generalized linear mixed-effects model on number of syllables in the non-final portion of the trisyllabic sequences, grouped by condition. Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Condition	**Estimate**	**SE**	**z-value**	**p-value**	**Sig.**
(Intercept)	1.37	0.414	3.30	0.0009	***
Normal	-2.62	0.31	-8.40	< 2e-16	***
Slow	-1.46	0.29	-5.13	2.95e-07	***
Very slow	-1.04	0.28	-3.69	0.000	***

4.2.2. Durations of the non-final part of the targeted location names

Figure 16 presents violin plots with overlaid boxplots illustrating the distribution of non-final part durations of the targeted location name in the recorded sentences across the four experimental conditions (i.e., "unswallowed", "normal", "slow" and "very slow"). Outliers were removed according to the 1.5 interquartile range (IQR) rule per condition. Corresponding descriptive statistics are provided in *Table 18*. The average duration is longest under the "unswallowed" condition (M = 0.360, SD = 0.070), followed by "very slow" (M = 0.353, SD = 0.089), "slow" (M = 0.302, SD = 0.066) and "normal" (M = 0.235, SD = 0.054).

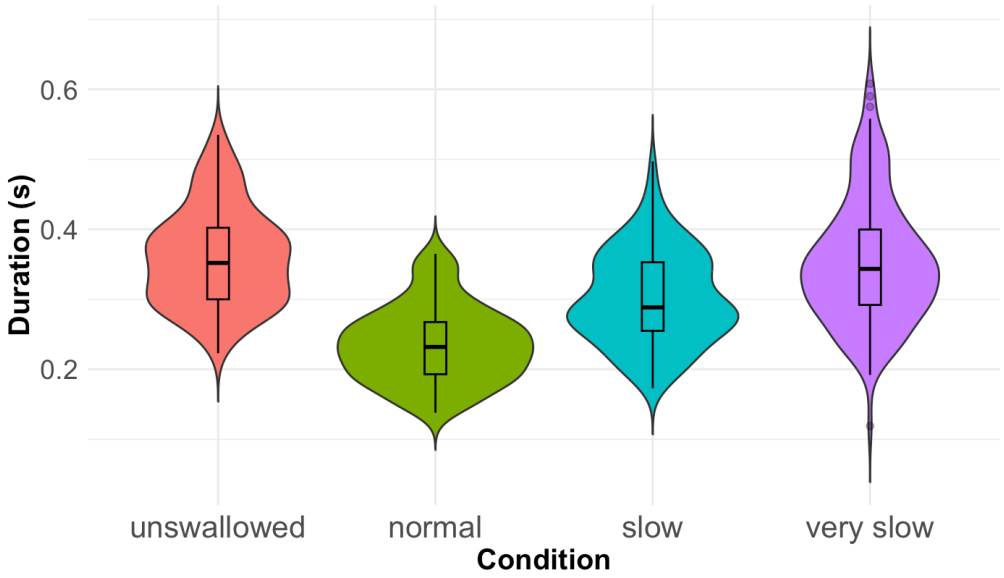

Figure 16. Violin plot with overlaid boxplot showing the distribution of non-final part durations of the targeted location names across the four conditions.

Table 18. Descriptive statistics of non-final part durations of the targeted location names (in seconds) by condition

Condition	Mean	SD	Min	Q1	Median	Q3	Max	n
Unswallowed	0.360	0.070	0.223	0.300	0.352	0.402	0.535	148
Normal	0.235	0.054	0.138	0.193	0.232	0.268	0.365	147
Slow	0.302	0.066	0.173	0.255	0.289	0.353	0.497	142
Very slow	0.353	0.089	0.119	0.292	0.344	0.400	0.608	142

A linear mixed-effects model was fitted to analyze the effect of condition on the durations of non-final-syllable portion of the trisyllabic sequences, with condition as the fixed effect. The default treatment coding scheme of the *lmertest* package was applied to the variable condition. Random intercepts and slopes for condition were specified for both participants and tokens. The *R* code for the model is provided in (16).

(16) model <- lmer (dur.non.fin ~ condition + (1 + condition | participant) + (1 + condition | token), data = df)

The results , as summarized in *Table 19*, showed that the "unswallowed" condition has significantly longer duration than the "normal" condition (Estimate = -0.12, SE = 0.01, t = -13.52, p < .001) and significantly longer duration than the "slow" condition (Estimate = -0.05, SE = 0.01, t =

-5.00, p <.001). In contrast, the “unswallowed” condition has shorter duration than the “very slow” condition， but the difference is not statistically significant (Estimate = 0.01, SE = 0.02, t = 0.28, p = 0.780).

Table 19. Summary of fixed effects in the linear mixed-effects model on non-final part duration. Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Condition	Estimate	SE	t-value	p-value	Sig.
(Intercept)	0.36	0.02	23.06	3.46e-15	***
Normal	-0.12	0.01	-13.52	1.72e-09	***
Slow	-0.05	0.01	-5.00	0.000	***
Very Slow	0.01	0.02	0.28	0.780	

5. Discussion

5.1. Summary of results and answering the research questions

Several acoustic patterns associated with retroflex-triggered “Beijing Swallowing” were identified in the trisyllabic production task. The results suggest that retroflex-triggered “swallowing” in native speakers of Beijing Mandarin involves merging the syllable that contains the retroflex and its preceding syllable into one while reducing their overall duration, as well as lowering the average F3 in the syllable rime preceding the retroflex. Comparison between the “swallowed” and “rhotacized” realizations suggest that, while syllable-final rhotacization and retroflex “swallowing” behave similarly in terms of syllable merger, duration reduction, and F3 lowering, syllable-final rhotacization decreases the average F3 of the rime even further than retroflex “swallowing” in Beijing Mandarin. This acoustic difference suggests that the two processes should not be treated as equivalent. Furthermore, the vowel following the swallowing-triggering retroflex was found to leave detectable acoustic trace in the F3 value of the rime preceding the retroflex after “swallowing”, suggesting that the rime following the retroflex is not fully dropped during this process.

The results from the sentence production task indicate that, while the frequency of Beijing speakers’ “swallowing” behavior is still influenced by speech rate, acoustic cues associated with

retroflex “swallowing” can be observed in slow, and even occasionally in very slow speech. This suggests that, although “Beijing Swallowing” is clearly not yet fully phonologized, it may be undergoing a process of phonologization, challenging Han’s view (2024) that the process is fully phonetic.

5.2. Limitations

Several limitations can be identified in the present study:

The participant group of the current study may not be sufficiently representative for a broader population, limiting the generalizability of the results. Critically, the sample size of 15 participants is too small compared to the whole Beijing Mandarin speaking population, which might be in millions[6]. Since all participants included in the final analysis were educated at college level or above, their speech may be more influenced by Standard Mandarin, as Standard Mandarin is predominantly used in education according China’s language policy (Ingebretson, 2025). The age distribution within the current sample group is also notably uneven. 11 out of the 15 participants were between 18 and 25 years old, with an overall mean age of just 28.6 and maximum age of 44. Consequently, the results might be more representative of the younger generation.

Since it is hard to construct minimal pairs that contrasts only in the rime of the middle syllable, the coarticulatory effect from the onset of the final syllable was not fully controlled for in the analysis on the effect of the vowel following the retroflex on the rime in the initial syllable. For instance, the sequences “杜十娘” |**tu^{51}. ʂə^{35}.** njɑŋ^{35}| and “读书人” |**tu^{35}. ʂu^{55}.** ʐən^{35}| were considered a comparable pair, as the non-final part (marked with bold and underline) contrasts only in the rime of the middle syllable in the underlying and “unswallowed” forms. However, in the “swallowed” and “rhotacized” forms (presumably realized as [**tuɻ^{51}** njɑŋ^{35}] and [**tuɻ^{35}.** ʐən^{35}] under both conditions), the onset of the final syllable, in this case, /n/ and /ʐ/, directly precedes the initial syllable’s rime. A study by Delvaux et al. (2002) on French nasal vowels found that nasalization has a notable effect on F3, not

[6] To the author’s knowledge, there are no existing data on the number of native Beijing speakers. However, considering the permanent residents in Beijing exceeded 20 million in 2023 (https://www.beijing.gov.cn/renwen/bjgk/rk/rktj/202403/t20240322_3597338.html), this estimation is unlikely to be over-exaggerated.

necessarily on formant value but formant amplitudes. Nonetheless, while the direction of the effect is not clear, this implies potential coarticulation not accounted for in the current design, which might affect the validity of the results.

The effect of word frequency was also not accounted for in the current design. Although the stimuli used in both parts of the experiment were generally familiar, high-frequency trisyllabic sequences, there were still differences in their usage frequency that could have influenced participants' "swallowing" behavior. For example, some sequences in the trisyllabic production task, such as "迪士高" |ti^{35}. ʂə^{51}. kɑʊ^{55}| ('disco') and "地主婆" |ti^{51}. ʈ͡ʂu^{214}. pʰɔ^{35}| ('landlady', typically used in historical contexts), may be perceived as more "old-fashioned" and could be more familiar to older speakers than to younger ones. As a result, older speakers might find these words easier to "swallow". Two of the participants (Participants 05 and 06) also reported after completing the experiment that they noticed their familiarity with certain location names significantly influenced their "swallowing" behavior during the sentence production task. Specifically, they noted being more likely to "swallow" a name if they frequently encountered and used it in its "swallowed" form in everyday speech.

In terms of methodological limitations, duration reduction may not be a good indicator for "swallowing" in the sentence production task, as speech rate itself fundamentally affects the duration of an utterance. Moreover, the method used for calculating the number of syllables in the non-final part of recorded utterances might be of limited validity. While the non-final part of surface "unswallowed" realizations was expected to be 100% disyllabic, only 87% were identified as disyllabic in the trisyllabic production task, and 74% in the sentence production task (see *Figure 7* & *Figure 15*). This might suggest that syllable merger at Window I (Wee, 2014, see *Figure 1* & *Figure 2*) could also occur in "unswallowed" speech, presumably due to the "medium-weak-strong" prosodic structure of Mandarin (Chao, 2005; Yan & Lin, 1988; Wang & Wang, 1993; as cited in Han, 2024). However, the possibility that the current method tends to underestimate the proportion of monosyllabic realizations should also be considered. Namely, some "swallowed" cases may have been identified as "unswallowed" in the sentence production task. One possible reason for this is

that the method was originally developed based on speech data from the trisyllabic production task of a single participant (Participant 01). Therefore, its generalizability to other participants and to the sentence production task is limited. Alternative syllable boundary detection methods, such as the "Mark regions by syllables…" function from the *Praat Vocal Toolkit* (Corretge, 2024), were also tested but found to have limited precision.

5.3. Is there actually phonologization going on?

Aside from the methodological limitations addressed above, the most important issue to discuss is how sufficiently the results of the current study prove the phonologization of retroflex-triggered "swallowing". While the results are not contradictory to the potential phonologization of the process, they cannot be taken as conclusive evidence of phonologization.

The method used to test the presence of phonologization, namely, the speech rate method proposed by Solé (1994), is of limited validity compared to other criteria established by other studies briefly discussed in *Section 2.3*. The acoustic features associated with retroflex "swallowing" identified in the current study, especially the consistent lowering of the F3 in the rime preceding the retroflex, could be linked to the criterion of categoricality / discreteness for phonologization (Shahin, 2011). Namely, if this F3 lowering pattern can be consistently found in the rime preceding a retroflex obstruent in a swallowing-eliciting context across different items, speakers, and contexts, this could suggest that retroflex "swallowing" is controlled by speakers on a more phonological level. Although participants' trisyllabic production in the current study exhibits certain patterns hinting at categoricality (i.e., consistent F3 drop), due to the highly controlled nature of the task, in which speakers are intentionally targeting "swallowing", the results could not be used as indicators of phonologization. Formant analysis was also not conducted for the more natural-speech like sentence production task. This is because the phonetic context surrounding the retroflex was not controlled for due to lexical constraints. Furthermore, item- and participant-specific variation should also be analyzed to examine whether retroflex "swallowing" meets the categoricality / discreteness criteria, which was beyond the scope of the current study.

Furthermore, Yu (2021) proposed an individual-difference perspective on phonologization. According to Yu, phonologization is not necessarily gradient and accumulative, but rather represents different speakers' linguistic knowledge. Namely, if some speakers consistently produce and control a phonetic variant, this can indicate phonologization is at least occurring for these individuals. Considering this, analyzing the individual variance in participants' swallowing realizations may shed some light of the investigation on phonologization. It was noticed during the experiment that some participants, especially younger ones, were less affected by the usage frequency of stimuli and maintained relatively stable "swallowing" realizations across utterances, while others reported finding less frequent words hard to "swallow". This, although being an informal observation, might reflect potential difference in these participants' knowledge of retroflex "swallowing". However, restricted by the timeframe, the individual-level analysis was not included in the current study.

The ultimate goal to achieve through analyzing the phonologization of "Beijing Swallowing", ideally, is to provide a possible theoretical framework to account for the conditioning factors behind the seemingly unsystematic "swallowing" patterns and sub-patterns identified in Han (2024). The author of the current thesis hypothesizes that, while "Beijing Swallowing" originates from a single phonetic process driven by speech rate and articulatory economy, it may have diverged into multiple processes over the past decades, reflecting different degrees of phonologization. This might explain the significance of retroflex "swallowing" among all "swallowing" processes, as retroflexion demands greater articulatory effort compared to other segments (e.g., Malghani et al, 2022), causing retroflex to be more frequently and consistently "swallowed", leading to a higher degree of phonologization.

5.4. Suggestions for future study

Based on the findings of the present study and the limitations addressed above, several suggestions can be proposed for future research on "Beijing Swallowing". Since the current study is limited to read speech, which may not be fully representative for Beijing speakers' natural "swallowing" behavior, it is suggested that further study targeting spontaneous speech should be

carried out. The investigation should also be expanded to other acoustic properties, for instance pitch, in order to capture the suprasegmental features of "swallowing" like its effect on tone realization, as well as "swallowing" triggered by other segments. With a longer timeframe, employing alternative methods for syllable boundary detection is recommended to improve the validity of the results, for instance, using intensity instead of glottal pulses for syllable boundary detection. Other more conclusive criterion for phonologization should also be employed, especially the categoricality / discreteness criteria. Moreover, item- and participant-specific variation in retroflex "swallowing" should also be analyzed, in light of Yu's (2021) individual-difference perspective on phonologization.

6. Conclusion

The current study provided a preliminarily acoustic profile of retroflex-triggered "Sound Swallowing" in familiar trisyllabic sequences in Beijing Mandarin. Several acoustic patterns associated with the process were identified, including merging the weak syllable that contains the swallowing-triggering retroflex onset and its preceding syllable into one and reducing their overall duration, while lowering the average F3 value of the rime preceding the retroflex. Furthermore, results suggest that the retroflex pattern reported in previous studies, namely, the syllable rime preceding the retroflex takes on a final [ɻ], and the syllable containing the retroflex gets fully elided, is not accurate. Rather, syllable-final rhotacization and retroflex "swallowing" in Beijing Mandarin should be treated as two distinct processes.

The investigation on the relationship between speech rate and the frequency of "swallowing" indicated that, although still sensitive to speech rate, acoustic characteristics associated with retroflex "swallowing" identified in the trisyllabic production task can be found in slow, and even occasionally in very slow speech. Although not conclusive, this suggests that retroflex-triggered "Beijing Swallowing" might be undergoing phonologization.

References

Astraxan, E. B., Zav'jalova, O. I., & Sofronov, M. V. (1985). *Dialekty i natsional'nyj jazyk v Kitae [Chinese dialects and the national language of China]*. Nauka.

Bao, Z. (1990). *On the nature of tone* [Doctoral Theses]. http://dspace.mit.edu/handle/1721.1/14143

Boersma, P., & Weenink, D. (2025). Praat: doing phonetics by computer. http://www.praat.org/

Chao, Y. R. (1930). A system of tone-Letters. *La Maitre Phonetique*, *45*, 24–47. Reprinted in 1980 in Fangyan, 2, 81-82.

Chao, Y. R. (2005). *汉语口语语法 [A Grammar of Spoken Chinese]* (S. Lü, Trans.). 商务印书馆. (Original work published 1968)

Chirkova, K., & Chen, Y. (2012). *Beijing Mandarin, the language of Beijing*. HAL. Retrieved August 20, 2012, from https://hal.science/hal-00724219

Corretge, R. (2024). Praat Vocal Toolkit. https://www.praatvocaltoolkit.com

Delvaux, V., Metens, T., Soquet, A. (2002) French nasal vowels: acoustic and articulatory properties. Proc. 7th International Conference on Spoken Language Processing (ICSLP 2002), 53-56, doi: 10.21437/ICSLP.2002-51

Dressler, W. (1976). Morphologization of phonological processes (are there distinct morphonological processes?). In A. Juilland (Ed.), *Linguistic Studies Presented to Joseph H. Greenberg* (pp. 313–337). Anma Libri.

Dressler, W. U. (1985). *Morphonology, the dynamics of derivation*. Ann Arbor: Karoma Publishers, Inc.

Handel, Z. (2017). The Sintic languages: phonology (2nd ed.). In G. Thurgood & R. J. LaPolla (Eds.), *The Sino-Tibetan Languages* (pp. 85–110). Routledge.

Hyman, L. M. (2008). Enlarging the scope of honologization. *UC Berkeley Phonology Lab Annual Reports*, *4*(4). https://doi.org/10.5070/p73zm91694

Hyman, L. M. (1975). *Phonology: Theory and analysis*. Holt, Rinehart & Winston. https://archive.org/details/hymanphonologytheoryandanalysis1975pdf

Ingebretson, B. (2025). "Speak standard Mandarin, write standard characters": Mandarin language promotion and its effect on minority languages in China. In W. Wei & J. Schnell (Eds.), *The Routledge Handbook of Endangered and Minority Languages* (pp. 213–227). Routledge.

Joseph, B. D., & Janda, R. D. (1988). The how and why of diachronic morphologization and demorphologization. In M. Hammon & M. Noonan (Eds.), *Theoretical Morphology* (pp. 193–210). Academic Press.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). LmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lee, W. (2005). *A phonetic study of the "Er-hua" rimes in Beijing Mandarin*. 9th European Conference on Speech Communication and Technology. https://doi.org/10.21437/interspeech.2005-433

Li, D. C. S. (2006). Chinese as a lingua franca in greater china. *Annual Review of Applied Linguistics*, *26*, 149–176. https://doi.org/10.1017/s0267190506000080

Li, Q., Chen, Y., & Xiong, Z. (2017). Tianjin Mandarin. *Journal of the International Phonetic Association*, *49*(1), 109–128. https://doi.org/10.1017/s0025100317000287

Liang, H. (2021, March 26). 大栅栏"的"栅"应该怎么读？（汉字里的故事）. *People's Daily Overseas Edition*, 11. https://paper.people.com.cn/rmrbhwb/html/2021-03/26/content_2040174.htm

Lín, M., Yán , J., & Sūn, G. (1987). Běijīnghuà liǎng zì zǔ zhèngcháng zhòngyīn de chūbù shíyàn [First look at the normal stress of bi-syllabic constituents in Beijing Mandarin]. *Fāngyán*, *1*, 57–73.

Lin, T. (1982). 北京话儿化韵个人读音差异问题 [The issue of individual variation in the pronunciation of rhoticized finals in Beijing Mandarin]. *语文研究*, *2*, 9–14.

Maini, E. (2020, May 31). *Interquartile range to detect outliers in data*. GeeksforGeeks. https://www.geeksforgeeks.org/machine-learning/interquartile-range-to-detect-outliers-in-data/

Malghani, F. A., Bano, S., & Veesar, Z. A. (2022). Is [+Back] feature enough to distinguish retroflex consonants from other coronals? *Webology*, *19*(2), 8159–8171.

McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). Montreal Forced Aligner: Trainable text-speech alignment using Kaldi. Proc. Interspeech 2017, 498-502. https://doi.org/10.21437/Interspeech.2017-1386

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Qualtrics. (2025). *Qualtrics XM Platform* [Computer software]. Qualtrics. https://www.qualtrics.com

R Core Team (2025). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Shahin, K. (2011). Acoustic testing for phonologization. *The Canadian Journal of Linguistics / La Revue Canadienne de Linguistique*, *56*(3), 321–343. https://doi.org/10.1353/cjl.2011.0026

Solé, M. (1994). New ways of analyzing sound change: Speech rate effects. *Belgian Journal of Linguistics, 9*(1), 21-44. https://doi.org/10.1075/bjl.9.03sol

Solé, M. J., & Ohala, J. J. (2010). What is and what is not under the control of the speaker: Intrinsic vowel duration. *Laboratory Phonology*, *10*. https://www.researchgate.net/publication/265000108_What_is_and_what_is_not_under_the_control_of_the_speaker_Intrinsic_vowel_duration

Tencent Holdings Limited. (2025). *WeChat*. Retrieved from https://www.wechat.com

Vennemann, T. (1972a). Phonetic analogy and conceptual analogy. In T. Vennemann & T. H. Wilbur (Eds.), *Schuchardt, the Neogrammarians, and the Transformational Theory of Phonological Change: Four Essays* (pp. 181–204). Athenäum.

Vennemann, T. (1972b). Rule inversion. *Lingua*, *29*, 209–242.

Wang, J., & Wang, L. (1993). 普通话多音节词时长分布模式 [Duration distribution pattern of Mandarin polysyllabic words]. *中国语文*, *2*, 112–116.

Wee, L. H. (2008). *Casual speech elision of tianjin trisyllabic sequences*. https://roa.rutgers.edu/files/931-1007/931-WEE-0-0.PDF

Wee, L. H. (2014). Casual speech elision and tone sandhi in Tianjin trisyllabic sequences. *International Journal of Chinese Linguistics*, *1*(1), 71–95. https://doi.org/10.1075/ijchl.1.1.03wee

Wee, L. H., Yan, X., & Lu, J. (2005). Tianjin fangyan de tunyin xianxiang [Swallowing sounds in Tianjin]. *Linguistic Sciences*, *17*, 66–75. http://journal15.magtechjournal.com/Jwk_yykx/CN/Y2005/V4/I4/66

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. Retrieved from https://ggplot2.tidyverse.org

Xing, K. (2021). *Phonetic and Phonological Perspectives On Rhoticity in Mandarin* [Doctoral Dissertation]. https://research.manchester.ac.uk/en/studentTheses/phonetic-and-phonological-perspectives-on-rhoticity-in-mandarin

Yan, J., & Lin, M. (1988). 北京话三字组重音的声学表现 [Acoustic properties of stress in trisyllabic sequences in Beijing Mandarin]. *方言 [Dialects]*, *3*, 227–237.

Zhang, Q. (2005). A Chinese yuppie in Beijing: Phonological variation and the construction of a new professional identity. *Language in Society*, *34*(03). https://doi.org/10.1017/s0047404505050153

Zhang, Q. (2021). Emergence of social meaning in sociolinguistic change. In L. Hall-Lew, E. Moore, & R. J. Podesva (Eds.), *Social Meaning and Linguistic Variation* (pp. 267–291). Cambridge University Press. https://doi.org/10.1017/9781108578684

Zhang, Y. (2014). A corpus based analysis of lexical richness of Beijing Mandarin speakers: variable identification and model construction. *Language Sciences*, *44*, 60–69. https://doi.org/10.1016/j.langsci.2013.12.003

Zhū, D. (1987). Xiàndài Hànyǔ yúfǎ yánjiū de duìxiàng shi shénme [What is the aim of modern Chinese grammar studies?]. *Zhōngguó Yǔwén*, *5*, 321–329.

Appendix 1: Demographic profile of participants

Participant	Age	Gender	Family background	Highest education	Knowledge in other Mandarin variants	Knowledge in other languages
01	20	Female	Two Beijing parents	Undergraduate	None	English
02	19	Female	No Beijing parents	Undergraduate	None	English
03	25	Male	One Beijing parent	Postgraduate	None	English
04	20	Female	One Beijing parent	Undergraduate	None	English, Korean
05	24	Male	Two Beijing parents	Postgraduate	None	English
06	24	Male	One Beijing parent	Postgraduate	Shandong dialect	English
07	18	Male	No Beijing parents	Undergraduate	None	English
08	42	Male	Two Beijing parents	Undergraduate	None	English
09	44	Female	Two Beijing parents	Undergraduate	Henan dialect	English
10	43	Male	Two Beijing parents	Undergraduate	Sichuan dialect	English
11	37	Female	Two Beijing parents	Undergraduate	None	English
12	44	Female	Two Beijing parents	Technical college	None	English
13[7]	79	Male	Two Beijing parents	Highschool	None	None
14	24	Female	Two Beijing parents	Postgraduate	None	English, Korean
15	20	Male	Two Beijing parents	Undergraduate	None	English, German
16	25	Male	Two Beijing parents	Undergraduate	None	English

[7] Excluded from analysis (see *Footnote 3*).

Appendix 2: Stimuli

- **Part 1: Trisyllabic production task**
    - Experimant stimuli:

	Stimulus	Gloss	Transcription	MoA of retroflex	Vowel before retroflex	Vowel after retroflex
1	八十九	'Eighty-nine'.	|ba^{55}. ʂɚ^{35}. tɕjoʊ^{214}|	fricative	/a/	/ɚ/
2	八只狗	'Eight dogs'	|ba^{55}. t͡ʂɚ^{55}. koʊ^{214}|	affricate	/a/	/ɚ/
3	大师级	'Master level'	|ta^{51}. ʂɚ^{55}. tɕi^{35}|	fricative	/a/	/ɚ/
4	打直球	'Cast a straight ball'	|ta^{214}. t͡ʂɚ^{35}. tɕʰjoʊ^{35}|	affricate	/a/	/ɚ/
5	西什库	Location name	|ɕi^{55}. ʂɚ^{35}. ku^{51}|	fricative	/i/	/ɚ/
6	西直门	Location name	|ɕi^{55}. t͡ʂɚ^{35}. mən^{35}|	affricate	/i/	/ɚ/
7	迪士高	'Disco'	|ti^{35}. ʂɚ^{51}. kɑʊ^{55}|	fricative	/i/	/ɚ/
8	地质锤	'Geological hammer'	|ti^{51}. t͡ʂɚ^{51}. t͡ʂʰweɪ^{35}|	affricate	/i/	/ɚ/
9	不识数	'Cannot count'	|pʰu^{51}. ʂɚ^{35}. ʂu^{51}|	fricative	/u/	/ɚ/
10	不知道	'Do not know'	|pu^{51}. t͡ʂɚ^{55}. tɑʊ^{51}|	affricate	/u/	/ɚ/
11	杜十娘	A fictional character	|tu^{51}. ʂɚ^{35}. njɑŋ^{35}|	fricative	/u/	/ɚ/
12	渎职罪	'Malfeasance'	|tu^{35}. t͡ʂɚ^{35}. tsweɪ^{51}|	affricate	/u/	/u/
13	巴蜀菜	'Sichuan cuisine'	|pa^{55}. ʂu^{214}. tsʰaɪ^{51}|	fricative	/a/	/u/
14	霸主级	'Dominant level'	|pa^{51}. t͡ʂu^{214}. tɕi^{35}|	affricate	/a/	/u/
15	大树根	'Big tree root'	|ta^{51}. ʂu^{51}. kən^{55}|	fricative	/a/	/u/
16	大主管	'Chief manager'	|ta^{51}. t͡ʂu^{214}. kwan^{214}|	affricate	/a/	/u/
17	洗漱包	'Toiletry bag'	|ɕi^{214}. ʂu^{51}. pɑʊ^{55}|	fricative	/i/	/u/
18	吸住它	'Suck it'	|ɕi^{55}. t͡ʂu^{51}. tʰa^{55}|	affricate	/i/	/u/
19	地鼠洞	'Groundhog burrow'	|ti^{51}. ʂu^{214}. tʊŋ^{51}|	fricative	/i/	/u/
20	地主婆	'Landlady'	|ti^{51}. t͡ʂu^{214}. pʰɔ^{35}|	affricate	/i/	/u/
21	不属于	'Does not belong to'	|pu^{51}. ʂu^{214}. y^{35}|	fricative	/u/	/u/
22	不注意	'Not paying attention'	|pu^{51}. t͡ʂu^{51}. i^{51}|	affricate	/u/	/u/
23	读书人	'Literate person'	|tu^{35}. ʂu^{55}. ʐən^{35}|	fricative	/u/	/u/
24	堵住它	'Block it'	|tu^{214}. t͡ʂu^{51}. tʰa^{55}|	affricate	/u/	/u/

    - Training stimuli:

	Stimuli	Gloss	Transcription
1	西客站	'Location name'	|ɕi^{55}. kɤ^{51}. ʈʂan^{51}|
2	电视台	'TV station (location name)'	|tjɛn^{51}. ʂɚ^{51}. tʰaɪ^{35}|
3	什刹海	'Location name'	|ʂɚ^{35}. ʈʂʰa^{51}. xaɪ^{214}|

- **Part 2: Sentence production task**

  - Experimant stimuli:

	Person name	**Transcription**	**Location name (targeted)**	**Transcription**
1	李红	|li^{214}. xʊŋ^{35}|	西什库	|ɕi^{55}. ʂɚ^{35}. ku^{51}|
2	王芳	|wɑŋ^{35}. fɑŋ^{55}|	西直门	|ɕi^{55}. t͡ʂɚ^{35}. mən^{35}|
3	张丽	|ʈʂɑŋ^{55}. li^{51}|	阜成门	|fu^{51}. ʈʂʰɤŋ^{35}. mən^{35}|
4	刘静	|ljoʊ^{35}. tɕɪŋ^{51}|	积水潭	|tɕi^{55}. ʂweɪ^{214}. tʰan^{35}|
5	赵玲	|ʈʂɑʊ^{51}. lɪŋ^{35}|	菜市口	|tsʰaɪ^{51}. ʂɚ^{51}. kʰoʊ^{214}|
6	李明	|li^{214}. mɪŋ^{35}|	美术馆	|meɪ^{214}. ʂu^{51}. kwan^{214}|
7	王强	|wɑŋ^{35}. tɕʰjɑŋ^{35}|	什刹海	|ʂɚ^{35}. ʈʂʰa^{51}. xaɪ^{214}|
8	张勇	|ʈʂɑŋ^{55}. jʊŋ^{214}|	德胜门	|tɤ^{35}. ʂɤŋ^{51}. mən^{35}|
9	刘伟	|ljoʊ^{35}. weɪ^{214}|	白石桥	|paɪ^{35}. ʂɚ^{35}. tɕʰjɑʊ^{35}|
10	赵刚	|ʈʂɑʊ^{51}. kɑŋ^{55}		灯市口	|tɤŋ^{55}. ʂɚ^{51}. kʰoʊ^{214}|

  - Training stimuli:

	Person name	**Transcription**	**Location name**	**Transcription**
1	李红	|li^{214}. xʊŋ^{35}|	西客站	|ɕi^{55}. kʰɤ^{51}. ʈʂan^{51}|
2	刘伟	|ljoʊ^{35}. weɪ^{214}|	电视台	|tjɛn^{51}. ʂɚ^{51}. tʰaɪ^{35}|
3	王芳	|wɑŋ^{35}. fɑŋ^{55}|	什刹海	|ʂɚ^{35}. ʈʂʰa^{51}. xaɪ^{214}|

Appendix 3: Descriptive statistics on average F3 values in initial syllable's rime

- Overall average F3 values of initial syllable's rime:

Gender	Phon.before	Context	Condition	Mean	SD	Min	Q1	Median	Q3	Max	n
Female	/a/	a_u	unswallowed	2521	162	2165	2405	2528	2646	2873	28
			swallowed	2355	199	2036	2240	2319	2507	2817	26
			rhotacized	2236	143	1955	2118	2235	2356	2495	56
		a_ɚ	unswallowed	2543	173	2108	2485	2552	2666	2879	28
			swallowed	2371	219	2017	2242	2368	2440	2865	26
			rhotacized	2236	143	1955	2118	2235	2356	2495	56
	/i/	i_u	unswallowed	3254	187	2833	3130	3220	3416	3573	28
			swallowed	2696	274	2169	2561	2699	2909	3267	27
			rhotacized	2289	153	1979	2175	2301	2405	2669	55
		i_ɚ	unswallowed	3264	205	2895	3139	3295	3390	3766	26
			swallowed	2669	265	2191	2538	2660	2833	3193	28
			rhotacized	2289	153	1979	2175	2301	2405	2669	55
	/u/	u_u	unswallowed	2780	198	2443	2627	2793	2870	3169	28
			swallowed	2438	331	1876	2179	2440	2690	3115	28
			rhotacized	2143	211	1702	1988	2096	2294	2630	56
		u_ɚ	unswallowed	2826	201	2522	2649	2802	2969	3165	28
			swallowed	2338	358	1826	2103	2222	2578	3154	28
			rhotacized	2143	211	1702	1988	2096	2294	2630	56
Male	/a/	a_u	unswallowed	2344	157	2072	2220	2332	2457	2635	32
			swallowed	2085	162	1788	1989	2020	2163	2440	29
			rhotacized	2050	135	1757	1963	2024	2135	2381	64
		a_ɚ	unswallowed	2306	133	2091	2209	2295	2397	2588	31
			swallowed	2106	170	1840	1984	2086	2214	2499	31
			rhotacized	2050	135	1757	1963	2024	2135	2381	64
	/i/	i_u	unswallowed	2834	128	2527	2788	2842	2881	3076	28
			swallowed	2210	178	1860	2070	2214	2325	2666	31
			rhotacized	1995	134	1690	1895	2003	2087	2314	61
		i_ɚ	unswallowed	2800	124	2524	2728	2811	2884	3017	29

			swallowed	2214	189	1870	2096	2176	2286	2676	32
			rhotacized	1995	134	1690	1895	2003	2087	2314	61
	/u/	u_u	unswallowed	2565	181	2334	2395	2549	2711	2957	32
			swallowed	2100	214	1697	1929	2095	2201	2588	32
			rhotacized	1885	188	1536	1737	1846	1977	2317	63
		u_ɚ	unswallowed	2554	205	2192	2408	2495	2705	3017	32
			swallowed	1947	211	1561	1777	1937	2131	2322	32
			rhotacized	1885	188	1536	1737	1846	1977	2317	63

- Overall average F3 values of initial syllable's rime in the 60-80% interval:

Gender	**Vowel before retroflex**	**Vowel after retroflex**	**Mean**	**SD**	**Min**	**Q1**	**Median**	**Q3**	**Max**	**n**
Female	/a/	/ɚ/	2139	242	1825	1977	2067	2247	2736	25
		/u/	2241	385	1703	1949	2188	2386	3353	26
	/i/	/ɚ/	2242	339	1594	2035	2235	2515	2829	27
		/u/	2481	257	2013	2358	2512	2562	2994	25
	/u/	/ɚ/	2109	318	1707	1892	2048	2238	3064	26
		/u/	2259	350	1687	2002	2190	2565	3138	27
Male	/a/	/ɚ/	1843	155	1638	1743	1815	1908	2217	30
		/u/	1802	166	1541	1699	1788	1903	2210	26
	/i/	/ɚ/	1939	105	1765	1856	1939	2021	2142	29
		/u/	2054	217	1635	1888	2004	2182	2620	31
	/u/	/ɚ/	1716	161	1309	1619	1684	1819	2142	30
		/u/	1990	293	1453	1799	1980	2201	2611	32

Appendix 4: Information brochure and consent form (with English translation)

研究信息告知书

研究项目：探究北京话“吞音”的音系化——来自卷舌音的声学证据

亲爱的参与者，

您好，您将参与的研究项目是《探究北京话“吞音”的音系化——来自卷舌音的声学证据》。该项目由阿姆斯特丹大学人文学院语言学系本科生朱语盈在导师 A.T. Benders 副教授的指导下进行。在研究开始之前，请您仔细阅读本告知书，了解相关实验流程与注意事项。

研究目的

本研究关注北京话中被称为“吞音”的语音现象，即在快速或随意的语流中某些音节似乎“消失”或被“吞掉”。本研究关注的是卷舌音（一类发音时舌头会向后卷的辅音，如汉语拼音中 *zh*、*ch*、*sh*）。

参与者要求

我们正在招募成年北京话母语者。参与者应在北京出生, 并在北京生活和接受教育至少至 18 岁。我们也需确保您没有任何可能影响您在实验中表现的已知身心障碍。在实验结束后，您将填写一份关于您的年龄、性别、家庭与语言背景、教育经历等信息的问卷。

实验流程

实验将在安静房间中进行。您将坐在桌前，面前放置一台 MacBook 笔记本，用于展示词语、句子与简单指令。实验期间将有实验员全程陪同，提供必要的口头指引。

整个实验包括两个部分，您将在实验部分被录音：

1. **词语/词组朗读**：您将朗读屏幕上展示的一系列词语或词组，分别包含“吞音”、“非吞音”与“儿化”三种形式。该部分约需 20-30 分钟。
2. **句子朗读**：您将以不同语速朗读屏幕上展示的一系列句子。该部分约需 15-20 分钟。

整个实验过程预计总时长在 1 小时以内。

自愿参与

本研究遵循完全自愿参与的原则，您在任何时候都可以中止参与，不会产生任何不良后果。如您决定在研究结果发表前退出，已收集的资料将会被永久删除；但若数据已经被匿名处理，则无法删除，因为无法追溯至个体。

风险与保险

本研究不会带来超出日常生活范围的风险。以往的类似研究表明，参与者基本不会感到不适或压力。阿姆斯特丹大学为所有研究活动统一提供责任保险。

数据的保密与处理

所收集的信息仅用于本项目研究，不会在公开场合披露您的任何个人信息。录音也不会被公开播放。研究数据将加密存

储，且与个人信息分开。只有研究人员能接触到这些信息。匿名化数据将保留十年，未匿名数据将在研究结束后尽快删除。

关于数据保护的知情权

您有权随时向研究人员索取关于您在欧盟《通用数据保护条例》（GDPR）下的相关信息与权利。

研究结果回馈

如您愿意，我们可以在研究完成后向您发送一份研究结果摘要。

联系方式

如需进一步了解项目内容，请联系：

- 朱语盈
  电话：+31 638970564
  邮箱：yuying.zhu@student.uva.nl
  地址：Spuistraat 134, 1012VB 阿姆斯特丹，荷兰
- A.T. Benders 副教授（指导教师）
  电话：+31 (0)205250000
  邮箱：a.t.benders@uva.nl
  地址：Spuistraat 134, 1012VB 阿姆斯特丹，荷兰

如您对本研究有任何投诉，可联系阿姆斯特丹大学人文学院伦理委员会秘书：

邮箱：commissie-ethiek-fgw@uva.nl

地址：Binnengasthuisstraat 9, 1012 ZA 阿姆斯特丹，荷兰

Information brochure for

Investigating the Phonologization of "Sound Swallowing" in Beijing Mandarin: Acoustic Evidence from Retroflex Segments

Dear participant,

You will be taking part in the research project *Investigating the Phonologization of "Sound Swallowing" in Beijing Mandarin: Acoustic Evidence from Retroflex Segments* conducted by Yuying Zhu under supervision of Dr. A.T. Benders at the University of Amsterdam, Faculty of Humanities. Before the research project can begin, it is important that you read about the procedures we will be applying. Make sure to read this brochure carefully.

Purpose of the research project

This project investigates a phenomenon in Beijing Mandarin known as "*Tunyin* (*Beijing Swallowing*)", where certain sounds seem to disappear or get "swallowed" in fast or casual speech. The focus is on a specific group of sounds that involve curling the tongue back (e.g., *zh*, *ch* and *sh* in Mandarin).

Who can take part in this research?

We are inviting adult native speakers of Beijing Mandarin to participate in this research. All participants should be born in Beijing and raised and educated there until at least the age of 18. We also need to make sure that you do not, to the best of your knowledge, have any physical or cognitive conditions that might affect your performances in the experimental tasks. After participating in the experiment, you will be asked to fill out a questionnaire about your age, gender, family background, language background, education, etc.

Instructions and procedure

During the experiment, you will be seated in front of a desk in a quiet room, with a MacBook laptop on the desk for displaying stimulus sequences or sentences as well as corresponding instructions. An experimenter will be present throughout the session to guide you and provide necessary verbal instructions.

The entire procedure will consist of two main components. You will be recorded in both parts of the experiment.

1. Word / sequence production task: You will be asked to read out a list of words or phrases displayed on the screen at varying conditions (i.e., "swallowed", "un-swallowed" or "rhotacized"). This will take about 20-30 minutes.
2. Sentence production task: You will be asked to read out a list of sentences displayed on the screen at varying speaking rates. This will take about 15-20 minutes.

The entire session is expected to last less than an hour.

Voluntary participation

You will be participating in this research project on a voluntary basis. This means you are free to stop taking part at any stage. This will not have any consequences and you will not be obliged to finish the procedures described above. You can always decide to withdraw your consent later on. If you decide to stop or withdraw your consent prior to publication of the research results, all the information gathered up until then will be permanently deleted. However, if information has been anonymised, it cannot be deleted because it is not possible to trace back the information to individual participants.

Discomfort, Risks & Insurance

The risks of participating in this research are no greater than in everyday situations at home. Previous experience in similar research has shown that no or hardly any discomfort is to be expected for participants.
For all research at the University of Amsterdam, a standard liability insurance applies.

Confidential treatment of your personal details

The information gathered over the course of this research will be used for the purpose of this research project. Your personal details will not be used in publications, and we guarantee that you will remain unidentifiable in all publications. Audio recordings will also never be shown in public.
The data gathered during the research will be encrypted and stored separately from the personal details. These personal details and the encryption key are only accessible to members of the research staff.
Anonymised data will be stored for a period of 10 years The non-anonymised data will only be stored as long as is necessary for the research and will be deleted as soon as possible.

Data subject rights according to the GDPR

Participants can request more information from the researcher at any time about their rights as data subjects under the EU privacy law, the GDPR.

Reimbursement

If you wish, we can send you a summary of the general research results at a later stage.

Further information

For further information on the research project, please contact Yuying Zhu (phone number: +31 638970564; email: yuying.zhu@student.uva.nl; Spuistraat 134, 1012VB Amsterdam, The Netherlands) and Dr. A.T. Benders (phone number: +31 (0)205250000; email: a.t.benders@uva.nl; Spuistraat 134, 1012VB Amsterdam, The Netherlands).
If you have any complaints regarding this research project, you can contact the secretary of the Ethics Committee of the Faculty of Humanities of the University of Amsterdam, commissie-ethiek-fgw@uva.nl; Binnengasthuisstraat 9, 1012 ZA Amsterdam, The Netherlands.

知情同意书

“本人特此声明，我已被清晰告知有关研究项目《探究北京话“吞音”的音系化——来自卷舌音的声学证据》的相关内容。该研究由阿姆斯特丹大学人文学院朱语盈在 A.T. Benders 副教授的指导下进行，相关信息已在信息告知书中详尽说明。我的所有疑问均已得到满意答复。

我知晓并同意在完全自愿的基础上参与本研究。我有权在任何时候撤回本同意书，而无需说明理由。我知晓自己可以在研究进行过程中随时终止参与，且即使研究结束后，我仍可撤回同意。如我选择中止或撤回同意，先前所收集的所有信息将被永久删除。

我理解，如研究结果被用于学术发表或以其他形式公开，所有内容都将经过匿名处理。在未获得我的明确许可之前，任何第三方不得查看我的个人信息。

若我现在或今后需要进一步了解研究详情，可联系朱语盈（电话：+31 638970564；邮箱：yuying.zhu@student.uva.nl；地址：Spuistraat 134, 1012VB 阿姆斯特丹，荷兰）或 A.T. Benders 副教授（电话：+31 (0)205250000；邮箱：a.t.benders@uva.nl；地址同上）。

如我对本研究有任何投诉，可联系阿姆斯特丹大学人文学院伦理委员会秘书：
邮箱：commissie-ethiek-fgw@uva.nl
地址：Binnengasthuisstraat 9, 1012 ZA 阿姆斯特丹，荷兰。

我同意：

- 参与本研究 ☐是 / ☐否
- 在实验过程中被录音 ☐是 / ☐否
- 我的个人信息将被储存十年 ☐是 / ☐否

此知情同意书一式两份：

…… ……. ……

参与者姓名 日期 签名

“我已就该研究进行了进一步说明，并声明愿意在我的能力范围内回答关于研究的任何问题。”

…… ……. ……

研究人员姓名 日期 签名

Informed consent form

'I hereby declare that I have been clearly informed about the research project *Investigating the Phonologization of "Sound Swallowing" in Beijing Mandarin: Acoustic Evidence from Retroflex Segments* at the University of Amsterdam, Faculty of Humanities, conducted by Yuying Zhu under supervision of Dr. A.T. Benders as described in the information brochure. My questions have been answered to my satisfaction.

I realise that participation in this research is on an entirely voluntary basis. I retain the right to revoke this consent without having to provide any reasons for my decision. I am aware that I am entitled to discontinue the research at any time, and that I can always withdraw my consent after the research has ended. If I decide to stop or withdraw my consent, all the information gathered up until then will be permanently deleted.

If my research results are used in scientific publications or made public in any other way, they will be fully anonymised. My personal information may not be viewed by third parties without my express permission.

If I need any further information on the research, now or in the future, I can contact Yuying Zhu (phone number: +31 638970564; email: yuying.zhu@student.uva.nl; Spuistraat 134, 1012VB Amsterdam, The Netherlands) and Dr. A.T. Benders (phone number: +31 (0)205250000; email: a.t.benders@uva.nl; Spuistraat 134, 1012VB Amsterdam, The Netherlands).

If I have any complaints regarding this research, I can contact the secretary of the Ethics Committee of the Faculty of Humanities of the University of Amsterdam; email: commissie-ethiek-fgw@uva.nl; Binnengasthuisstraat 9, 1012 ZA Amsterdam, The Netherlands.

I consent to:

- participate in this research ☐yes / ☐no
- audio recordings being made ☐yes / ☐no
- my personal details to be stored for a period of 10 years ☐yes / ☐no

Signed in duplicate:

…… ……

Name participant Date Signature

'I have explained the research in further detail. I hereby declare my willingness to answer any further questions on the research to the best of my ability.'

…… ……

Name researcher Date Signature